(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 December 2000 (28.12.2000)

PCT

(10) International Publication Number
WO 00/78934 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/GB00/02423

(22) International Filing Date: 22 June 2000 (22.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

9914589.8	22 June 1999 (22.06.1999)	GB
0008162.0	3 April 2000 (03.04.2000)	GB

(71) Applicant *(for all designated States except US)*: **SCHOOL OF PHARMACY, UNIVERSITY OF LONDON** [GB/GB]; 29-39 Brunswick Square, London WC1N 1AX (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **NIZETIC, Dean** [HR/GB]; School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX (GB). **GROET, Jürgen** [DE/GB]; School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX (GB).

(74) Agent: GILL JENNINGS & EVERY; Broadgate House, 7 Eldon Street, London EC2M 7LH (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/78934 A2

(54) Title: DIAGNOSIS AND TREATMENT OF ALZHEIMER'S DISEASE

(57) Abstract: A gene has been identified in human chromosomes, on human chromosome 21 in the region q11-q21. The gene has been sequenced and shown to have sequence homologies with the ubiquitin specific protease family. The recombinant gene has been cloned into *E.coli*, and the product shown to have ubiquitin specific protease properties. It is believed that the activity of the enzyme may be involved in the pathway leading to neurofibrillar tangles observed in Alzheimer's disease and/or in general neurotoxicity leading to progressive neuronal degeneration and cell death. The invention relates to processes for diagnosing Alzheimer's disease, for treating Alzheimer's disease and for investigating the mechanism of Alzheimer's disease. The sequence of the enzyme has GenBank accession No. AF 134213. The gene has the name USP25, approved by the HUGO Nomenclature Committee.

Diagnosis and Treatment of Alzheimer's Disease

The present invention relates to Alzheimer's Disease, a possible pathogenesis of the disease, processes for investigation and diagnosis of the disease, and for treatment of the disease. In particular the invention relates to the discovery of a gene, and the activity of its product, on human chromosome 21.

Background of the Invention

Alzheimer's disease (AD)

Alzheimer's disease (AD) (Alzheimer A.) is the most common form of dementia in humans, affecting up to 5% of human population above the age of 65 (Terry RD *et al.)*. Less than 15% of all AD are a familial, autosomally inherited form, and more than half of these are explained by causative mutations in presenilin gene (PS-1) on chromosome 14. A minority of familial cases are found to carry causative mutations in presenilin 2 (PS-2) gene on chromosome 1 and amyloid precursor protein (APP) on chromosome 21. The majority of AD cases (>85%) are known as sporadic AD, although the distinction between familial and sporadic AD becomes very elusive in the late onset AD (age of onset >70), due to an insufficient number of testable living relatives. It is therefore possible that the sporadic AD, specially the most prevalent late onset type, is also an unrecognised genetic disease.

The main pathohistological findings that define a diagnosis of AD are two kinds of neuropathological protein deposits in AD brains, in particular in the hippocampus, neocortex and amygdala: neurofibrillary tangles (NFT) and amyloid plaques (AP). Amyloid plaques are extracellular accumulations of amyloid material which occur in normal ageing brains (senile plaques), but are much more prominent in AD, and do not occur in other neurodegenerative dementias. Neurofibrillary tangles are large

intraneuronal, usually perikaryal masses that are never observed in neurons from normal brains of any age, but are only visible in neurons of brains suffering from neurodegenerative diseases including AD.

All individuals with Down syndrome (DS) develop early AD changes in brains

Trisomy 21 (DS) individuals are known to be at risk of developing AD in middle life, though the clinical picture may in some cases be obscured by the pre-existing mental retardation, characteristic of DS. The brains of DS individuals in virtually 100% of cases exhibit neuropathology indistinguishable from AD by the end of the fourth decade of life. This includes NFTs and their major ultrastructural component, the PHFs, as well as amyloid plaques. Immunohistochemical studies have shown that amyloid is deposited in brains of trisomy 21 individuals about 50 years before that seen in the normal ageing population. In normal brains, soluble $A\beta^{\infty}$ is detectable at low levels. These levels increase with ageing, but do not necessarily lead to AD-like pathology. The highly insoluble $A\beta^{\alpha}$ is detected in brains of ageing euploid individuals only with the full AD neuropathology, and clinical AD. However, in DS brains (Teller), $A\beta^{\alpha}$ preceeds $A\beta^{\infty}$ but is somehow kept soluble until the third and fourth decade of life, when AD neuropathology develops in its full picture, including NFTs and APs.

Overexpression of *β*AP produces amyloid plaques, but not NFT, in AD model animals

Amyloid β-peptide is known to play a major role in the pathogenesis of AD. Aβ is derived by proteolytic processing from the β-amyloid precursor protein (β APP), encoded by a gene located on human chromosome 21q21.2.

A number of studies have attempted generating a mouse model with typical pathohistological changes of AD. The most abundant are approaches by over-expressing the APP gene product. Constructs driven by powerful promotors have been intruduced into mouse genome leading to over-expression of the normal form of APP holoprotein (refs 4-10 in Games), the carboxy-terminal peptides of APP, including the part from which $A\beta^{40}$ (AP) is carved (Shoji), or of the APP genes bearing well-established mutations corresponding to those causing some types of familial early onset AD (FEOAD) (Games, Calhoun). The study by Games succeeded in producing widespread deposits of βAP resembling amyloid plaques of AD, but in addition of over-expressing the mutated form of APP, this construct was 18 fold more over-expressed than any previous attempts due to the choice of promotor. The attempt by Calhoun generated plaques as well as evidence of widespread neuronal loss in the hippocampus and neocortex of transgenic mice. There was no evidence whatsoever in any of the above mentioned models of the presence of PHF or NFT, or of staining with antibodies against any epitopes characteristic of PHF or NFT in the transgenic mice. The outcome of these attempts strongly proofs that overexpression of APP alone is not sufficient to explain the whole of AD the histopathology, which nevertheless occurs in 100% of trisomy 21 cases. The obvious imposing explanation is that the elevated dose of other chromosome 21 encoded gene(s) must play a role, perhaps together with that of APP.

The only mouse neuronal cells in which the presence of epitopes characteristic of PHFs and NFTs (such as τ -NFT and ubiquitin), in addition to staining with β AP, was demonstrated (Richards), are neurons transplanted into normal mouse recipients, originating from developing brains of mice trisomic for the whole of mouse chromosome 16, which in some of its portion is syntenic to most of the long arm of human chromosome 21 (containing DNA analogous to human regions from 21q11 through to 21q22.1). The donor mice however die in utero due to severe malformations of several organ systems, and are an imperfect model for DS, because of the

trisomy of portions of mouse chromosome 16 syntenic also to other human chromosomes, and because of the lack the trisomy of the segments syntenic to the last telomeric portions of human chromosome 21 (21q22.2-qter). Though highly disputed, and not reproducible in all mouse strains (Lane *et al* 1996), the study by Richards hinted at the possibility that the trisomy of the human segment 21q11-21q22.1 (which includes APP) could be all that is necessary for the generation of both types of major AD hallmarks: PHF (NFT) and amyloid plaques. However, when mice Ts65DN were generated (Reeves) trisomic for the two thirds of this segment (from and including APP to the end of the segment in 21q22.1), neither amyloid nor PHF (NFT) staining or pathology was observed, despite an increased dose of APP mRNA as demonstrated by Northern blot hybridisation in these mice, which live to adulthood. This result suggests that trisomy of genes in the most proximal part of the 21q (21q11-q21) could play a central role in causing AD pathology in DS, both in generating the NFT (PHF), and perhaps in combination with APP overexpression, in generating the βAP deposits and plaques.

Paired Helical Filaments (PHF) are aggregates of ubiquitin (Ub) and altered Tau (τ)

The progressive formation of 10-nm thick filaments wound into a helix with a half-periodicity of 80 nm (thus termed paired helical filaments or PHFs) is one of the most typical and pathognomonic pathohistological features characterizing AD. The PHFs accumulate as large perikaryal masses called neurofibrillary tangles (NFTs), or occur as small bundles in dystrophic neurites that form plaques in AD brain. Biochemical analysis of the core material of the PHF, through analysis of pronase resistant core of the filament and a proteinaceous smear by SDS-PAGE and immunoblotting revealed that the PHF is composed of only two proteins: ubiquitin and Tau (τ).

Several independent reports found NFTs and PHFs strongly and specifically staining with antibodies against ubiquitin (Mori, Perry, Manetto). One report studied inclusion bodies characteristic of other neurodegenerative diseases (Manetto) and found ubiquitin strongly present not only in PHFs of AD, but also in Pick bodies of Pick disease, Lewy bodies of Parkinson disease and NFTs of progressive supranuclear palsy (PSP). However, ubiquitin was not present in Hirano bodies and granovacuolar degeneration, indicating that although present in a wide variety of intracellular inclusion bodies of degenerating neurons, ubiquitin was not indiscriminatly present in all neuronal inclusions.

Ubiquitin (Ub) is a 76 amino acid polypeptide present in all eukaryotic cells, and highly conserved in evolution. Ubiquitin is conjugated to a target protein through an isopeptide bond between the ε-amino group of Lys in a target protein (ubiquitination), a process mediated by three groups of enzymes: ubiquitin activating enzymes (E1), ubiquitin conjugating enzymes (E2) and ubiquitin ligases (E3). Ubiquitinated proteins exist in a monoubiquitinated form, or a multiubiquitin chain: the former is not a degradation signal, while the latter, Lys-48-linked ubiquitin-ubiquitin(n) conjugate, works as a strong degradation signal when joined to a Lys in a target protein. Protein conjugated to polyubiquitin is then very rapidly and efficiently degraded by non-lysosomal, ATP-dependent degradation by the 26-S proteasome. Two classes of enzymes termed ubiquitin specific proteases (USPs) and ubiquitin-C-terminal hydrolases (UCHs) are capable of de-conjugating the ubiquitin-ubiquitin and ubiquitin-protein links, thereby converting polyubiquitin into mono-ubiquitin, and de-coupling (deubiquitinating) ubiquitin from the target protein, with the result of preventing the degradation of the target protein.

The class of UCH enzymes tends to include relatively small proteins (about 35-40kD) which have low specificity for ubiquitinated proteins. BAP1 is a UCH but is unusual in that it is larger, having a weight of nearly 100kD.

Several such enzymes are known, the sequences of which show some sequence homologies especially in two domains, the Cys and His domains.

Very broadly, two main functions have been observed among the various members of the USP (UBP) superfamily (Wilkinson 1997, Wilkinson and Hochstrasser 1998). The first is the generation of free ubiquitin from precursor fusion proteins or from peptide-linked polyubiquitin after proteolysis of the targetted protein, and the second is de-ubiquitination. When a protein is targetted for ubiquitin mediated degradation, it is linked to ubiquitin via an isopeptide bond between the C-terminus of ubiquitin and a lysine ε-amino group(s) of the acceptor protein. Once the conjugate is formed, it can have only two fates: non-lysosomal proteolysis mediated by the 26S proteasome resulting in total protein degradation, or de-conjugation from ubiquitin (de-ubiquitination), resulting in the rescue of the target protein from degradation (Wilkinson 1997, Wilkinson and Hochstrasser 1998).

The PHF-s are strongly staining for a sub-fragment of τ, the τ-NFT. This abberrant form of τ has an enormously high affinity for τ itself, much higher than the physiological affinity for microtubules. Therefore, once the threshold concentration of the τ-NFT is reached, aggregates leading to PHF become inevitable, and their structural assembly process is relatively well explained. What is not clear, is what causes the initial generation of the τ-NFT.

Each single molecule of the amino-end cleaved τ-NFT is ubiquitinated at the Lys residue with a single Ub molecule. Ubiquitination with a single Ub is an insufficient signal to trigger the 26S proteasome degradation, which normally requires polyubiquitinated target protein. USP's are capable of hydrolysing the thiol bonds between individual ubiquitins and reversing poly-Ub back to mono-Ub. The same enzymes also cleave the single Ub off the substrate, recycling both the substrate and Ub.

Screening of cDNA libraries prepared from frontal cortex of an AD patient and from fetal brain revealed a microtubule-associated protein τ (Goedert). In PHFs, an altered form of τ was detected, which was processed

at the amino-terminal end, leaving only the carboxy-terminal third of the polypeptide (Wischik). This altered form, termed PHF-τ was found to bind the insoluble core of the PHF very strongly, more then 20 times stronger than the physiological afinity of full length τ for microtubules (Wischik, Beyreuther). It was further revealed that ubiquitin present within PHFs in conjugated to amino-terminally processed τ in the mono-ubiquitinated form predominantly, with the conjugation sites localized to the microtubule-binding region (Morishima-Kawashima). The same authors suggested that the ubiquitin-degradation system could be affected in the degenerating neurons of AD brain.

E6AP (Human Papilloma Virus E6 associated protein) functions as one of the two so far detected ubiquitin ligases (attaching ubiquitin and labelling for degradation) for the master tumour suppressor p53 (Scheffner et al. 1993). Haupt et al. 1997, Kubbutat et al. 1997, Lane 1998).

Lack of functional E6AP accelerates the polyglutamine-induced neuronal cell death in the mouse model for the neurodegenerative disease Spinocerebellar-ataxia 1 (SCA1) (Cummings et al. 1999). Lack of E6AP gene in a mouse expressing the polyglutamine stretch mutation of SCA1 protein dramatically reduces the presence of ubiquitinated intranuclear neuronal inclusions, but drastically accelerates the neuronal degeneration and cell death (Cummings et al. 1999). A very similar effect has been observed in Huntington's Disease (HD), where a dominant negative mutant of a ubiquitin conjugating enzyme (UBC3), when co-expressed in cultured neurons with the huntingtin protein bearing the polyglutamine extension (mutation causing HD), drastically reduces the presence of ubiquitinated intranuclear neuronal inclusions, but drastically accelerates the neuronal degeneration and cell death (Saudou et al. 1998). If USP25 de-ubiquitinates a similar set of target proteins to the ones ubiquitinated by E6AP, then the overexpression of USP25 may lead to similar effects as the inhibition of ubiquitin conjugation by E6AP.

Lowe *et al.* (1990) show that a ubiquitin carboxyl-terminal hydrolase is selectively present in inclusion bodies characteristics of human neurodegenerative diseases. The authors have used immunohistochemical methods to determine the presence of the UCH. They found immuno reactivity in Lewy bodies but not in NFT's of AD.

In recent years a number of other protein modifying polypeptide tags have been identified. Many of these are related to ubiquitin and have high levels of identity and similarity (determined using the BLAST algorithm, for instance) to ubiquitin itself. There is a recognised super family of such proteins which have been termed ubiquitin-like proteins (UbL) (Gong et al. 1997, Schwarz et al. 1998). The yeast Smt3 and human SUMO-1 (PIC1, Sentrin, hSmt3C), SUMO-2 (hSmt3A) and SUMO-3 (hSMT3B) belong to the same family of UbL proteins with approximately 50% identity between themselves, and some 15-30% identity and 40-60% similarity in amino acid sequence to ubiquitin (Lapenta et al. 1997, Mannen et al. 1996, Kamitani et al. 1998, Saitoh and Hinchey, 2000). Yeast and human UBC9 are capable of conjugating equally yeast or human UbL-s, but not ubiquitin (Schwarz et al. 1998). The SUMO-1,-2 and -3 have the C-terminal glycine, necessary for ubiquitination of the target protein's lysine residue, but unlike ubiquitin, do not have the Lys48 residue necessary for the formation of polyubiquitin chains through isopeptide bonds, which are the signal for the proteasome degradation (Saitoh and Hinchey, 2000). Nevertheless, yeast Smt3 protein can rescue the mutant Mif2 phenotype, a deficient centromere binding protein resulting in chromosome missegregation (Meluh and Koshland 1995). SUMO-1, as well as SUMO-3 (and probably also SUMO-2) are all capable of being attached by UBC9 to RanGap1, a Ran GTP-ase activating protein (Kamitani et al. 1998). This ATP-dependent attachment is essential for the binding between modified RanGap1 and RanBP2 binding protein, in order to form functional nuclear pore complex, which controls export and import of molecules through the nuclear envelope (Mahajan et al. 1997, Matunis et al. 1998, Lee et al. 1998). In addition, UbL small proteins have

been shown to modify the death domains of Fas (Okura et al. 1996), Tumour necrosis factor receptor1 (Okura et al. 1996), PML (a tumour suppressor implicated in the pathogenesis of acute promyelocytic leukaemia) (Kamitani et al. 1998b) and Rad51/52 DNA repair proteins (Shen et al. 1996a). Their conjugating enzyme, UBC9, has been shown to interact by Y2H technique with RAD51/52 DNA repair proteins, and the master tumour suppressor p53 (Shen et al. 1996b). Another UbL is NEDD8 (Kamitani *et al* 1997).

UbL's are conjugated and cleared from their targets by enzymes. Several UbL hydrolase enzymes have been identified which convert precursor UbL to active UbL. Some such enzymes interact with ubiquitin itself as well as with other UbL's. Proteases involved in cleavage of conjugates of UbL with target protein have been identified for instance SENP1 and SUSP-1, which were recently cloned (Kim et al. 2000, Gong et al. 2000a), and found to specifically cleave SUMO-1,-2 and -3, but not ubiquitin and NEDD8. The first human enzyme with classical USP structure (Cys, His domains) for which dual specificity to both ubiquitin and ubiquitin like protein was demonstrated was very recently published USP21 on chromosome 1q21 (Gong et al. 2000b). However, opposite from SENP1 and SUSP-1, this enzyme cleaves ubiquitin and Nedd8, but not SUMO-1,-2 or -3 (Gong et al. 2000b).

The proximal third of the chomosome 21 long arm is an exceptionally gene-poor region of the human genome as estimated by a number of criteria (Shimizu et al., 1995, Yaspo et al., 1995, Gardiner 1996) the estimates of gene-density range from one gene in a megabase to one gene in six megabases of genomic DNA. Until recently only three full length genes had been mapped in this region: STCH (a member of the hsp70 family) (Brodsky et al., 1995), RIP140 (protein functionally interacting with a variety of nuclear receptors such as estrogen receptor), (Cavailles et al., 1995) and ANA (a previously mentioned member of the Tob/BTG1 family of tumour suppressors), (Kohno et al., 1998). This region is also an example of extremely highly methylated regions in the human genome.

Groet *et al.* (1998) describe a high-resolution bacterial contig map of 3.4 Mb of genomic DNA in human chromosome 21q11-q21, encompassing the region of elevated disomic homozygosity in Down's syndrome - associated abnormal myelopoiesis and leukemia, and which has shown a strong association with Alzheimer's disease. It was suggested that the high resolution bacterial clone overlap map should be the basis for deriving a more complete transcriptional map of that region of the chromosome. It was hoped that this would lead to an explanation of the chromosome 21q11 linkage in FEOAD families. In particular it was suggested that a modifier gene in that region could act together with the PSEN 1 gene to generate or modify the AD phenotype.

Valero *et al* 1999 have, in parallel but published after the first priority date hereof, identified the gene in human chromosome 21 as well as mouse chromosome 16. The gene products have close sequence homologies and are identical in the Cys, QQD and the regions. The authors suggest a role of the gene in AD. Valero *et al* name the gene USP25.

Further work by the present inventors has revealed and sequenced a new gene in the proximal third of chromosome 21 and has shown that the product of this gene has sequence homologies to USP's as well as ubiquitin specific protease properties. It is postulated that the gene product and USPs generally may have a role in AD. This further work has been published as Groet *et al* 2000 after the first priority date hereof.

According to the invention there is provided a new use of ubiquitin specific protease, an inhibitor thereof or the gene therefor, or a specific inhibitor or promoter of such a gene, in the manufacture of the composition for use in the treatment, diagnosis or prophylaxis of AD.

According to a further aspect there is provided a new use of ubiquitin-like specific protease, an inhibitor thereof or the gene therefor, or a specific inhibitor or promoter of such a gene, in the manufacture of the composition for use in the treatment, diagnosis or prophylaxis of AD.

There is also provided an *in vitro* method of diagnosis of AD in a human or animal in which a ubiquitin specific protease, or the gene therefor is detected in a sample taken from a patient suspected of suffering from AD, wherein an alteration in the level of expression or in the sequence of the translated SP is detected. In a further aspect there is also provided an *in vitro* method of diagnosis of AD in a human or animal in which a ubiquitin-like specific protease, an inhibitor thereof or the gene therefor, or a specific inhibitor or promoter of such a gene, in the manufacture of the composition for use in the treatment, diagnosis or prophylaxis of AD.

In the diagnostic method, the specific protease (SP) may be detected immunochemically, using antibodies specifically directed to the SP, or by contacting the sample with a substrate for SP under conditions whereby proteolytic reaction should take place, and detecting the product of enzymic reaction. The product of the reaction of the reaction may be detected photometrically (for instance fluorometrically) or immunochemically, for instance using an antibody to the cleavage product. The antibody may either react specifically with the cleavage product but not with the substrate, or else may react with both, in which case the cleavage product and the substrate must be capable of being separated from one an other for instance using gel electrophoretic techniques. Methods for the detection of changes in the DNA sequence or in methylation or in the quantity or sequence of mRNA, for instance using PCR techniques may be used in the diagnostic method, for instance carried out on the whole organism or on individual tissue sections or samples or cells, or several different cells in paralllel

The present invention also includes a method for investigating the pathogenesis of AD by the use of a USP.

In a further aspect of the invention there is provided a method for investigating the pathogenesis of AD by the use of a ubiquitin-like specific protease.

There is also provided in the present invention a method of synthesizing a protease having Cys, QQD and His domains specified in

sequence ID's Nos. 2, 3 and 4, respectively in which nucleic acid encoding the protease is introduced into a microorganism or a cell-line in a form in which it can be replicated, transcribed and translated, and the microorganism or cell-line is cultured under conditions whereby the nucleic acid is replicated, transcribed and translated to form protease, and the protease is collected.

Preferably in the invention the gene encoding the USP has been derived from the respective exons, usually mainly or wholly without the introns, of the naturally occurring gene from animal, preferably human, DNA, preferably from human chromosome 21 or the equivalent non-human animal chromosome, such as mouse chromosome 16.

Preferably the gene is incorporated into a construct, which preferably comprises ds DNA. Alternatively the construct may comprise ss DNA or RNA.

In the present invention, the ubiquitin specific protease or ubiquitin-like specific protease is preferably defined as a protein having ubiquitin specific protease or ubiquitin-like specific protease activity and including the following three domains having the specified sequences of the general formulae I, II and III

$$GXX^{1}\ NX^{2}GX^{3}TCX^{4}X^{5}X^{6}X^{7}X^{8}X^{9}QX^{10}X^{11} \quad \text{I}$$

$$QX^{12}DX^{13}X^{14}EX^{15}X^{16}X^{17}X^{18}X^{19}X^{20}X^{21}X^{22}X^{23}X^{24}X^{25} \quad \text{II}$$

$$YX^{26}LX^{27}X^{28}X^{29}X^{30}X^{31}HX^{32}GX^{33}X^{34}X^{35}X^{36}X^{37}GHYX^{38}X^{39}X^{40} \quad \text{III}$$

in which

X is the residue of a non-polar amino acid

X^{1} is the residue of an amino acid with an uncharged or basic R group

X^{2} is the residue of an amino acid with an uncharged R group

X^{3} is the residue of an amino acid with an uncharged R group

X^{4} is the residue of an amino acid having a relatively large uncharged R group

X^{5} is the residue of an amino acid having an uncharged R group

X^6 is the residue of an amino acid having a relatively polar uncharged R group

X^7 is the residue of an amino acid having an uncharged R group

X^8 is the residue of an amino acid having a non-polar R group

X^9 is the residue of an amino acid having an uncharged R group

X^{10} is the residue of an amino acid having an uncharged R group

X^{11} Is the residue of an amino acid having an uncharged R group

X^{12} is Q or H

X^{13} is the residue of an amino acid having a polar R group

X^{14} is the residue of an amino acid having an uncharged or basic R group

X^{15} is the residue of an amino acid having an uncharged R group

X^{16} is the residue of an amino acid having an uncharged R group

X^{17} is the residue of an amino acid having an uncharged polar R group or a basic R group

X^{18} is the residue of an amino acid having an uncharged or a basic R group

X^{19} is the residue of an amino acid having a polar R group

X^{20} is the residue of an amino acid having an uncharged R group

X^{21} is the residue of an amino acid having an uncharged polar R group or an acidic R group

X^{22} is the residue of an amino acid having an uncharged polar R group or a basic R group

X^{23} is the residue of an amino acid having a non-polar R group

X^{24} is the reside of an amino acid having an uncharged polar R group or an acidic R group

X^{25} is the residue of any amino acid

X^{26} is the residue of any amino acid

X^{27} is the residue of an uncharged or a basic amino acid

X^{28} is the residue of an amino acid having a non-polar R group

X^{29} is the residue of an amino acid having a non-polar R group

X^{30} is the residue of an amino acid having a non-polar R group

X^{31} is the residue of an amino acid having an uncharged R group

X^{32} is the residue of an amino acid having an uncharged polar R group or an acidic R group

X^{33} is the residue of an amino acid having an uncharged or an acidic R group

X^{34} is the residue of an amino acid having an uncharged R group

X^{35} is residue of an amino acid having an uncharged or a basic R group

X^{36} is the residue of an amino acid having an uncharged or a basic R group

X^{37} is the a bond or the residue of an amino acid having an uncharged polar R group or a basic R group, or is a bond

X^{38} is the residue of an amino acid having an uncharged R group

X^{39} is the residue of an amino acid having an uncharged R group and

X^{40} is the residue of an amino acid having an uncharged R group.

The first domain mentioned above is also known as the Cys domain, the second is also known as the QQD domain and the third is also known as the His domain. In the present invention the R group of an amino acid is the side chain attached to the α-carbon atom.

In the present invention, the term amino acid having a non-polar R group includes A, V, L, I, P, F, W, M and G. The term amino acid residue having an uncharged polar R group includes P, W, G, S, T, C, Y, N and Q. An amino acid having an acidic R group is selected from D and E. An amino acid residue having a basic R group is selected from K, R and H.

X is preferably F or most preferably L.

X^1 preferably has a basic R group, most preferably being K or R. Polar uncharged groups X^1 are preferably selected from Y and M. It is less preferred that X^1 is an amino acid residue with a polar R group.

X^2 is preferably a residue of an amino acid having a non-polar R group. Preferably it is selected from V, S, U, A, L and F.

In X^3, the R group is preferably a low molecular weight group, the X^3 residue preferably being selected from A and N.

X^4 is preferably a relatively high molecular weight residue, most preferably selected from Y and W.

X^5 is preferably selected from F, M, L and C.

X^6 is preferably a relatively low molecular weight residue, most preferably selected from N and S.

X^7 is preferably selected from A, C and S.

X^8 and X^9 are preferably selected from residues in which the R group is a C_{3-4} - alkyl group, most preferably selected from V, L and I. X^9 may alternatively be S.

X^{10} is preferably selected from the same groups as X^2. Most preferably it is selected from S, V, Q, A and T.

X^{11} is preferably selected from Y and, most preferably L.

X^{12} is preferably Q.

X^{13} is preferably selected from residues in which the R group is a C_1 or $_3$ - alkyl and is thus preferably V or A, most preferably V.

X^{14} is preferably selected from S, T, Q, L and H;

X^{15} is preferably selected from F, L, M and V.

X^{16} is preferably selected from T, L, N, F and C;

X^{17} is preferably selected from H, T, R, N and Q;

X^{18} is preferably selected from K, L, I, V, S, C and Y;

X^{19} is preferably the residue of an amino acid having a $C_{3\text{ or }4}$ - alkyl group as R, most preferably selected from L, I and V.

X^{20} is preferably L.

X^{21} is preferably the residue of an amino acid having an acidic R group, and is most preferably D.

X^{22} is preferably W.

X^{23} is preferably selected from the same groups as X^{19}. Most preferably it is L.

X^{24} is preferably a polar uncharged amino acid or an acidic amino acid residue, most preferably being selected from E, Q, D and A.

X^{25} is preferably an acidic residue, most preferably being D.

X^{26} is preferably selected from R, E, V, N, G, D and I; and

X^{27} is preferably selected from H, Y, V or F most preferably being H.

All of X^{28} to X^{31} are preferably selected from non-polar amino acids, most preferably in which the R group is hydrogen or a $C_{1 \text{ or } 4}$ alkyl group. X^{28} preferably has a small R group and is most preferably A or G. X^{29} and X^{30} are preferably selected from the same groups as X^{19}. X^{31} may be one of the same groups as X^{19} or may be M. Preferably X^{31} is V.

X^{32} is preferably selected from E, S and C.

X^{33} preferably has a polar R group or an acidic R group. Polar R groups are preferably Q or S.

In X^{34}, the R group is preferably non-polar, for instance being selected from the same groups as X^{19}, A and G.

X^{35} is preferably a polar R group which is uncharged, most preferably being selected from N, H, S and P, most preferably N.

X^{36} is selected from A, R, N, T, G and S;

R^{37} is preferably a bond.

R^{38} is preferably non-polar, most preferably being selected from Y, W, T, I and V. Most preferably it is W.

X^{39} is preferably a group in which R is polar, and is most preferably selected from A, S, T and V.

X^{40} is preferably selected from Y, L, S and I.

The SP should preferably have at least 200 residues, most preferably at least 500 residues, for instance between 750 and 1500 residues, most preferably around 1000 residues.

The preferred SP has Cys, QQD and His domains with the sequences of Sequence ID's nos. 2-4, respectively.

The preferred SP has the sequence ID no. 1.

The sequence homology of the protein having sequence ID no. 1 (USP25) with other USPs' is expressed in figure 3 hereof. The sequences of the USP's are derived from references mentioned above.

The SP is generally made by genetic engineering techniques. The nucleic acid in the microorganism used to synthesize the SP may be cDNA derived from mRNA, from human or animal sources. For instance the nucleic acid may be constructed from PACs which are publicly available, together with cDNA by suitable restriction enzyme cleavage, ligation and/or chain extension techniques. Alternatively the gene may be synthesized through other genetic engineering techniques.

The nucleic acid coding for a SP should encode the protein having the domains specified above. Preferably the nucleic acid is DNA having the portions of sequence ID no.5 which code for the respective domains or a sequence encoding the same protein sequence. The DNA sequence may include the regulatory regions (Untranslated regions, UTRs) from the native gene and may include $(GCC)_n$ repeats in the 5' regulatory, or transcribed and or translated region.

There is also provided in the present invention a non-human animal model to be used to investigate the mechanism of AD involving USP. The animal model may have the USP gene knocked out, hemi- or homozygously, or alternatively may be polysomic for the relevant gene. Alternatively the gene may be modified by disrupting the transcriptional control whereby the animal has reduced or increased levels of USP.

The USP of particular relevance to AD is believed to be that found on by the present inventors on human chromosome 21 long arm at q11-q21. It is believed that the gene is likely to be found in the analagous portion of mouse chromosome 16. The model animal is likely to be a mouse.

The invention provides also a nucleic acid construct comprising a USP gene encoding a product having (Cys, QQD and His regions as specified in sequence ID's Nos 2,3 and 4) and an origin of replication. Preferably the gene encodes sequence ID No. 1 or No. 5, or a USP-active

fragment or homologue thereof. A homologue is the corresponding gene from a non-human mammal. Most preferably the gene has bases 199-3367 of sequence ID No. 6, (represented as sequence ID No 7). A further aspect of the invention provides a nucleic acid construct encoding a product having Cys, QQD and His regions as defined in the general formulae I, II and III, respectively, and an origin of replication. A further aspect of the invention provides a nucleic acid construct encoding a product having Cys, QQD and His regions of sequence ID's 2, 3 and 4, respectively, or sequences having at least 20% identity with the specified sequences, preferably at least 50% identity, and an origin of replication. According to a further aspect of the invention there is provided a nucleic acid construct encoding a product having a contiguous sequence of at least 10 residues having at least 50% identity with a contiguous sequence of residues of sequence ID 1, and an origin of replication. In this aspect the product should preferably have a contiguous sequence of residues of sequence ID 1. Additionally or alternatively the product may have two or more, for instance three or four, contiguous sequences of at least 10 bases having at least 5% identity with respective contiguous sequences of sequence ID 1. In each case the level of similarity in the product with sequence ID 1 is preferably at least 50% (as determined using the BLAST algorithm), more preferably at least 70% in the related regions. The said contiguous sequences of sequence ID 1 may be the Cys, QQD and/or His regions thereof (sequence ID's 2, 3 and 4) or may be other regions for instance regions associated with the recognition of protein-ubiquitin or protein-ubiquitin-like protein targets.

Preferably the construct is formed from a vector having an origin of replication and the gene of interest. Preferably the construct also has a promoter for the initiation of transcription, generally derived from the vector. There is also provided a transformed microorganism or a transfected cell-line containing the construct wherein the construct is capable of being replicated. The cell-line may, for instance, be based on neurones,

neuroblasts or neuro epithelial cells and may be particularly useful for investigating the steps which take place in the generation of NFT's.

In a further aspect of the invention there is provided a protein product which is not a functional protease which has a contiguous sequence of at least 10 residues having at least 50% identity with a contiguous sequence of sequence ID 1. Preferably the product has a contiguous sequence of at least 20 residues, for instance 50 or more, having at least 20%, preferably at least 50%, identity with a contiguous sequence of sequence ID 1. The said contiguous sequences of sequence ID 1 may be the Cys, QQD and/or His regions thereof (sequence ID's 2, 3 and 4) or may be other regions, for instance regions associated with the recognition of protein-ubiquitin or protein-ubiquitin-like protein targets. In this aspect the level of similarity of the product with sequence ID 1 is preferably at least 50% (as determined using BLAST algorithm) more preferably at least 70% in the related regions. Preferably the product has contiguous sequences with at least 50% identity, preferably more than 90% identity, with the Cys, QQD and His regions of sequence ID 1 (sequence ID's 2, 3 and 4). Preferably the product has contiguous sequence which differ from those of sequence ID's 2, 3 and/or 4 in respect of one, two or three residues, preferably including the Cys residue which is the 9th residue of sequence ID 2, and/or one or more of the QQD residues which are the 1st to 3rd residues of sequence ID 3, and/or the His residue which is the 17th residue of sequence ID 4, most preferably the Cys residue of sequence ID 3. Preferably this embodiment of the product has sequence identity in the remaining residues with a contiguous sequence of at least 20% of sequence ID 1, preferably at least 50%, more preferably at least 80%, of sequence ID 1.

The novel protein products of the invention may be competitive inhibitors for the native protein product of sequence ID 1 and thus have useful therapeutic activity, for instance in the treatment of AD. Alternatively they may be useful components of kits for use in methods for investigating the activity of the protein product of sequence ID 1, or for identifying the

presence or level of the protein product of sequence ID 1 in a biological sample, for instance in a diagnostic test for AD.

The present invention further provides a protein having sequence ID 1 in substantially isolated form. In this context, "substantially isolated" means that the protein has been recovered from its source, whether that be a natural source, for instance from human tissue, or recombinant microorganism or transformed cell line source, in a process in which at least some of the components of the source have been removed from the protein. Preferably the protein is an active USP or ubiquitin-like specific protease.

It is believed that the ubiquitin specific protease activity of the protein having sequence ID 1 is responsible for its implication in the pathogenesis of AD. USP activity may be determined using the technique described in the Examples below, in which using bacteria cotransformed with the USP gene and with a reporter gene encoding a fusion protein which is a ubiquitin-conjugated detectable protein. The protein may be an enzyme detectable by direct enzyme reaction, by enzyme-linked immune assay techniques, by autoradio-graphically or by direct staining after gel separation under conditions suitable to separate ubiquitin and cleaved protein from fusion protein.

From experiments conducted to determine with which proteins the product having sequence ID 1 interacts, we have found that there is interaction with ubiquitin, polyubiquitin and various ubiquitin precursors, as well as HHR23A (Matsutani *et al* 1994, GenBank Accession No D21235) there is also interaction with other ubiquitin precursors, there is also interaction with other ubiquitin-like proteins and with proteins which are known to interact with ubiquitin-like proteins such as Sumo-3 (Mannen *et al* 1996, Kamitani *et al* 1996, Saitoh *et al* 2000, GenBank Accession No. NM 006937) and ubiquitin-like-specific conjugating enzyme 9 (Schwarz *et al* 1998, Lee *et al* 1998, GenBank Accession No. U 66867). The novel isolated protein and the novel protein having no protease activity may therefore be characterised further by having a positive interaction in a yeast-two-hybrid

procedure with one or more, preferably all three, proteins having the sequences of GenBank Accession Nos. D 21235, NM 006937 and U 66867. Ubiquitin-like specific protease activity may be determined using techniques analogous to those used to determine ubiquitin specific protease activity, by using a substrate which is a fusion protein of the ubiquitin-like protein of interest and a detectable protein, and using the usual separation and immune based or autoradiographic identification techniques.

In the present invention useful inhibitors of ubiquitin specific protease, for instance for use in the manufacture of a composition for treating AD, are ubiquitin analogues which compete with the ubiquitinated substrate and/or react with the protease enzyme so as to inactivate it. Known inhibitors include ubiquitin aldehyde, which has a C-terminal aldehyde group instead of a carboxylic group. Fragments of ubiquitin without activatable C terminal glycine, may also be inhibitors. Ubiquitin-like molecules, fragments thereof and C-terminal modified versions thereof may competitively inhibit USPs. The inhibitor may be a specific inhibitor and comprise a fusion protein for instance with a ubiquitin or UbL component and a recognition protein component for specific binding to the USP of interest.

The new model of pathogenesis of AD which the present inventors propose assumes an aberrant increase of the cleavage by USP, particularly USP25 as the key event triggering the creation of PHF-s, and hence the key event leading to the generation of NFT-s and the pathogenesis of AD, and other neurodegenerative conditions characterized by the presence of neurons with NFT-s or similar ubiquitinated inclusion bodies. This increase in activity would increase the rate of reduction of poly-Ub to mono-Ub, and increase the de-ubiquitination of the substrate. This would prevent the otherwise extremely rapid degradation of τ -NFT down the 26S proteasome route, and cause temporary increase in the intraneuronal concentration of τ -NFT. Since the τ -NFT has an abnormally high affinity for τ instead of microtubules, this would cause the seeding of PHF-s. Once the τ -NFT is seeded into the initial structure leading to PHF, it becomes resistant to both

ubiquitination and de-ubiquitination, perhaps by being sterically hindered and inaccessible to enzymes. This would block the removal of the last single Ub moiety from the τ -NFT, hence the presence of mono-Ub τ -NFT in PHF. The actual generation of τ -NFT from normal τ remains a mystery: it could either be the second hit necessary for the PHF seeding, or be part of normal neuron physiology, which is under normal conditions rapidly and efficiently polyubiquitinated and removed by the 26S proteasome. It is also possible that full length τ itself is physiologically efficiently polyubiquitinated and removed by the 26S proteasome, but an aberrant increase in USP activity as the primary event causes improper (partial) degradation of τ, by stopping the normal degradation process in its first steps, through an unusually quick conversion of polyubiquitinated to a monoubiquitinated substrate. The resulting partially cleaved, mono-ubiquitinated form could be the τ -NFT.

Presenilin-1 also uses the 26S proteasome pathway for its rapid and efficient degradation (Fraser). This degradation is deemed to be extremely rapid since ubiquitinated forms of presenilin have not been detected, and yet polyubiquitination is an absolute pre-requisite for a protein to be "labelled" for the 26S proteasome degradation. One of the most important functions of presenilin, pertinent to the pathogenesis of AD is that it appears to activate and promote the γ- secretase cleavage of APP C-terminal fragment which has already been generated (Borchelt). This γ- secretase cleavage is the last step necessary for the generation of βAP, the key component of amyloid plaques. An aberrantly increased activity of USP could slow-down the physiologically rapid degradation of presenilin, causing the increase in its concentration and activity, producing the similar effect as some of the mutations in pre-senilin causing the FEOAD. Indeed, studies on Presenilin-1 knockout mouse models (deStrooper) support the explanation that clinical FEOAD mutations in Presenilin-1 result in a gain of the function of PS1, and are not a haplotype insufficiency.

The known USP-s in the human genome are UBH on 5q33, FAFX

on Xp11, HAUSP-related gene on 3p21, HAUSP on 16p13 and USP-1 on 1q32. Related UCH genes are BAP-1 and UCH-L1 on 4p13. These and other USP's are reviewed in Baker *et al* which lists the GenBank Accession Nos. USP's identified since the publication of Baker *et al* are USP20, GenBank Accession No. NM 006676.1, and USP21 (Gong *et al* 2000b) GenBank Accession No. AF 233442. UbL specific proteases SUSP-1 and SENP-1 have GenBank accession nos. AF 196304 and AF 149770, respectively. All these specific proteases may be useful in the respective aspects of the invention mentioned above. We now add a new gene on chromosome 21- USP25, located in 21q11-q21. A genetic change leading to over-expression of any and all of these could trigger the AD pathogenesis. This does not rule out the alteration in pre-senilin function or the APP cleavage as alternative primary causes. This hypothesis just postulates that sporadic AD could be caused by the de-regulation of some of these enzymes. Both enhancement of function and/or copy number of some and inhibition of function and/or copy number of other USPs or UCHs could be the cause of the physiological disturbance. It is possible, in particular among USPs, that some members of the superfamily are more specialized for certain substrates. An overexpression of a competing, but less efficient "cousin" in the superfamily could then bring to the same effect as the underfunctioning of the specialist USP, which could be cause by a mutation. A mutation causing some familial cases of Parkinson's disease (PD) has been found in UCH-L1. In this case, the twofold reduction in the enzyme activity (caused by a mutation) causes PD. UCH-L1 does not stain AD-NFTs, but it does stain Lewy bodies in PD, and it can cause familial PD. Therefore, this enzyme may not be the primary candidate to cause AD.

USP-25 is located in a highly methylated chromosomal region, and the CpG island that occupies the 5' regulatory sequences and 5'UTR of USP25 is differentially methylated in a tissue specific fashion. The methylation could be the key mechanism through which the precise spatial and temporal regulation of USP expression in finely regulated. The breakdown in this

regulation could be age related. It could occur as a somatic change, only in affected neurons. On the other hand it maybe a constitutionally inherited allele, if it is true that the majority of SAD are actually unrecognized FAD with a late onset. If the theory of neuronal mosaicism for trisomy 21 (Potter) is proven correct, it would put USP-25 increased gene dose, just as in DS, in the driving seat causing the generation of NFTs, with or without a significant additive effect of the trisomy of APP. Other events affecting the 5' regulatory sequences could play a role: we have identified a (GCC)n repeat in the 5' regulatory region of USP25. Such repeats are known to frequently expand in population, and even somatically between cells one adult organism. Such repeat expansions are a cause of many genetic diseases, including myotonic dystrophy, fragile X syndrome and Huntington's disease.

The invention is illustrated further in the accompanying drawings in which:

Figure 1 represents a map of human chromosome 21

Figure 2 represents a map showing the location of exon trapped products from the experiments reported below

Figure 3 represents sequence homologies of USPs

and

Figure 4 represents the results of the experiment illustrating USP properties reported below.

The following specific description describes the work which has been carried out.

Examples

Tumour Samples

We identify a portion of human chromosome 21 homozygously deleted in non-small cell non carcinoma (NSCLC) for further study. The region contained the DNA marker with the highest NSCLC-associated loss of homozygosity (LOH), reported by Kohno *et al.* We found a shared region of overlap (SRO) for the hemizygous loss in other NSCLC. The current work is

to identify genes in the SRO which have a potential role in tumour suppression.

A total of 42 fresh NSCLC cases have been analyzed from the Croatian Tumour Bank (CTB), an initiative with set rules and criteria for accumulation of fresh clinical tumor specimens for molecular studies (Spaventi *et al.*, 1994). Of these, half (including tumors #47 and #61) were samples that were recently studied for LOH of the *NM23-H1* gene (Bosnar *et al.*, 1997), and the other half were fresh tumors (data obtainable from CTB, which also lists the tumor stage in the TNM system, grade, size and survival data). In each case, the tumors and normal lung tissue specimens (as evaluated by the surgeon) were frozen in liquid nitrogen in the operating room and further stored at -70°C. Genomic DNA was isolated using standard procedures (Sambrook *et al.*, 1989). For each sample, 4μm serial frozen sections were cut, mounted on glass slides, and stained with hematoxylin-eosin (H&E). A pathologist confirmed the histologic type of the tumor and evaluated the percentage of normal cells within the tumor. Only samples with less than 20% non-tumor cells were used in this study.

Markers and LOH Analysis

Microsatellite analysis was performed using polymerase chain reaction (PCR) with appropriate primer pairs (sequences and PCR conditions as in Genome Data Base, Johns Hopkins University, Baltimore, MD), where the forward primer only from each pair was 5' fluorescently labeled with Applied Biosystems (ABI; Foster City, CA) Big Dyes™ (6-FAM or HEX). Amplification products were analysed using an ABI 310 Genetic Analyzer. Size standards (GeneScan 350) were mixed with every sample for accurate sizing; the separation of the mixture of denatured fragments was achieved by electrophoresis through a 47 cm capillary (module GS STR POP4 C) for approximately 30 min. Raw data were analyzed using GeneScan and Genotyper software. LOH ratios were calculated exactly as described in the GeneScan Applications manual provided by ABI. For each individual allele's

fluorescence level, an average of 3 independent electrophoresis-analysis cycles on the ABI 310 was used for calculation.

Fluorescence In Situ Hybridization (FISH)

Unstained 4µm-thick paraffin block sections were fixed to glass slides, and a standard pretreatment protocol was followed for formalin-fixed, paraffin-embedded slides. P1-derived artificial chromosome (PAC)DNA was labelled with digoxygenin-11-dUTP (Boehringer Mannhein, Germany). Approximately 0.5 µg of each labelled PAC DNA sample was mixed with 5 µg of Cot1 DNA (Gibco BRL, Gaithersburg, MD), precipitated, denatured, allowed to preanneal, and then applied to a denatured slide and hybridized overnight. Slides were washed and signal detected using anti-digoxygenin-rhodamine, followed by DAPI counterstain. Images were captured using a Zeiss Axioskop microscope equipped with a charge-coupled device (CCD) Photometrics, Tucson, AZ) connected to an Apple Powermac 8100 computer. Images were captured on 3 levels of focus, and each level was examined for signals using SmartCapture software (Vysis, Inc., Chicago, IL). Only nuclei with signals were counted in each level, and the number of signals in each cell was determined. B: FISH using a pool of PACs 90B5, 126N20 and BAC 39112 as a probe on the paraffin embedded sections of the tumour #61. Two signal nuclei are predominant. C: FISH using a pool of PACs 73M5 and 135E14 as a probe on the paraffin embedded sections of the tumour #61. Single signal nuclei are predominant.

Northern Blot Analysis

The cDNAs were labelled by random priming and hybridized to human multiple tissue Northern blots (Clontech, Palo Alto, CA) containing 2 µg polyA + RNA per lane using the protocol recommended by the manufacturer. The exposure was for 14 hr to Molecular Dynamics (Sunnyvale, CA) Phosphorimager screens. The I.M.A.G.E. Consortium (Lennon *et al.*, 1996) cDNA clone ID 824710 and the Unigene clone A002B43 have been used as labelled probes in separate experiments.

Cleavage Analysis of Ubiquitin-Met-β-Galactosidase Fusion Protein

This analysis was performed essentially as described (Everett *et al.*, 1997). Model fusion protein ubiquitin-Met-β-galactosidase in a pACYC184 (Cm^r replicon) was represented by the plasmid pACYC-Ub-Met-β-gal, a kind gift of R. Everett. Plasmid pRB105 containing a *Saccharomyces cerevisiae* ubiquitin-specific protease UBP2 in an IPTG-inducible pBR322 (Amp^r replicon) was a kind gift of R. Baker, and was used as a positive control. The new gene *USP25* was cloned from nucleotide position 203 to nucleotide position 3367 (numbering as in GenBank AF 134213) into *SacI*/*SalI* cloning sites of the IPTG-inducible *Escherichia coli* expression vector pQE30 (Qiagen, Chatsworth, CA). The *E. coli* XL-1 blue cells were transformed using a standard rubidium chloride-heat shock method with the combination of pACYC-Ub-Met-β-gal and either pQE30 vector, pQE30-*USP25*, or pRB105, and each of the 3 cotransormants was selected on medium containing chloramphenicol (42 µg/ml) and carbenicillin (75 µg/ml). Western blots were prepared by electrotransfer to a nitrocellulose membrane (Schleicher & Schull, Keene, NH). The β-galactosidase-containing bands were detected by an anti-β-galactosidase polyclonal rabit antiserum (a kind gift of R. Everett) using an enhanced chemiluminescence (ECL) assay kit (rpn 2132; Amersham, Arlington Heights, IL) under conditions recommended by the manufacturer.

Exon -Trapping

DNA from the 2 PACs (73M5 and 1035 E14) was digested using *Bam*HI and *Bgl*II to completion or partially using *Sau* 3AI and exon trapping was performed with the resulting fragments in pLSP3B vector using standard technology.

Identification and cloning of USP25

Twelve sequenced exon-trapped products, when analysed using BLAST-N against public sequence databases, revealed clusters of overlapping cDNA clones. Sequences of our exon-trapped products matched

exactly the sequences of the cDNAs forming contigs with a large open reading frame (ORF). In three cases (see Fig.2): from EST 824710 to AA209364, from AA307805 to AA081200 and from N92952r to Z45010, our trapped exon sequences served to bridge the gaps in the gene sequence using PCR and suitable restriction, ligation and chain extension techniques. The combined sequence (sequence ID no.5 and GenBank accession number AF134213) revealed a 199 bp 5'UTR, start codon, an ORF of 3165 nucleotides encoding a protein of 1055 amino acids, a stop codon, a 3'UTR of 435 nucleotides and a polyadenylation signal. The total length (without the polyA) assembled is 3803 nucleotides. On multiple human tissue Northern blots (Fig.3) a band of 4.1 kb is visible in all 16 tissues tested (including the normal human lung tissue) with a varying intensity. It is most prominent in skeletal muscle and testis, and the latter tissue also reveals a prominent shorter hybridising transcript of 1.4 kb. All tissues also show a larger weaker band of 4.9 kb, which could be due to an alternative polyadenylation site.

In the course of this analysis, the whole genomic sequence of the two PACs (73M5 and 135E14) became publicly available by the German Human Genome Sequencing Consortium (EMBL accession numbers AJ010597 and AJ010598). Comparison of the genomic sequence with the overlapping cDNA clones and exon sequences revealed that 12 out of the 24 exons had been exon-trapped (hatched rectangles in Fig.2). It also became apparent that the region immediately preceeding the first exon of the gene, comprises the known chromosome 21 CpG island at D21S382 (also known as LL56 Not I linking clone on the Not I physical map of 21q, Ichikawa et al., 1993).

When the deduced polypeptide sequence was compared to Swissprot and other public databases using BLAST-P, a clear pattern of significant homologies ($e=10^{-6}$ to 10^{-29}) to proteins across the evolutionary spectrum of eukaryotes was found (Fig.4): all of these proteins belonged to the superfamily of ubiquitin specific proteases (USP-s) or ubiquitin carboxy-terminal hydrolases (UCH-s) (Baker et al., 1992, Swanson et al., 1996, Everett et al., 1997, Wilkinson 1997, Hansen-Hagge et al., 1998, Jensen et

al., 1998, Fujiwara et al., 1998, Wilkinson and Hochstrasser 1998, The *C. elegans* Sequencing Consortium 1998). The polypeptide sequences were most highly conserved around the three domains (the Cys box, the QQD box and the His box, Fig.4) known to be essential for the main function of these enzymes: the cleavage of ubiquitin at its carboxy terminus from extension proteins (ubiquitin precursors) and ubiquitinated proteins and protein fragments targetted for the degradation by the 26S proteasome pathway (Wilkinson and Hochstrasser 1998). The Cysteine residue at position 178 and the Histidine residues at positions 599 and 607 (marked with an asterix in Fig.4), which were shown to be an absolute requirement for the function of USP-s and UCH-s (Amerik et al., 1997, Hansen-Hagge et al., 1998, Wilkinson and Hochstrasser 1998) were found in the correct positions in the sequence of the new gene. The name of the protein USP25 has been approved by HUGO Nomenclature Committee.

The novel protein (USP25) cleaves ubiquitin from carboxy-terminal fusion proteins

The ability of USP25 to cleave a model ubiquitin fusion protein substrate was investigated by co-expression in *E. coli*. The complete coding sequence of USP25 was cloned into a T5-driven, IPTG inducible expression vector (pQE30). The new gene USP 25 was cloned from nucleotide position 203 to nucleotide position 3367 (numbering as in sequence ID no. 5 into *Sac I/Sal* cloning sites of the IPTG-inducible *E.coli* expression vector. As a positive control, the plasmid pRB105 containing a UBP2 gene encoding a *S.Cerevisiae* ubiquitin specific protease in an IPTG inducible and Amp^R vector was used. The XL-1 blue strain of *E. coli* was co-transformed with the plasmid containing a ubiquitin-Met-β-galactosidase model fusion protein in an IPTG-inducible and chloramphenicol resistant vector, in addition to either pQE30 vector, pQE30-USP25 or the positive control (pRB105). (Each of the 3 co-transformants was selected on medium containing chloramphenicol (42 µg/ml) and carbenicillin (75 µg/ml). Co-transformants were grown to exponential phase, IPTG induced, and the crude protein extracts from these

cultures were analysed by Western blot using an anti β-galactosidase antibody. (The western blots were prepared by electro transfer to a nitro cellulose membrane (Schleicher and Schuel.)). The β-galactosidase containing bands were detected by an anti-β galactosidase polyclonal rabbit anti serum using enhanced-chemiluminescence assay kit (ECL, Amersham rpn2132) under conditions recommended by the manufacturer.

As can be seen in Fig.4, the uncleaved Ub-Met-β-gal substrate (band labelled with an asterix in Fig.4, lane 4) converts to an 8 kDa shorter band (triangle in Fig.4) in the cells co-transformed with either USP25(lanes 5,6) or the yeast UBP2 expressing plasmid (lanes 9,10). Constitutive expression of USP25 (lane 5) is quite sufficient to cleave to completeness the low levels of model substrate. The more prominent and highly induced band migrating slightly further in the gel than the de-ubiquitinated cleavage product is the truncated form of β-galactosidase expressed by the XL-1 blue bacteria (compare to lanes 1,2 in Fig.4). This result demonstrates that the novel gene product named USP25 can efficiently function as a de-ubiquitinating enzyme.

From the homologies in the functional domains and from its ability to hydrolyse the bond between the C-terminal double glycine of ubiquitin and the linking methionine residue (Fig.5), it can be concluded that USP25 is a member of ubiquitin specific proteases.

We identify a portion of human chromosome 21 homozygously deleted in non-small cell non carcinoma (NSCLC) for further study. The region contained the DNA marker with the highest NSCLC-associated loss of homozygosity (LOH), reported by Kohno *et al.* We found a shared region of overlap (SRO) for the hemizygous loss in other NSCLC. The current work is to identify genes in the SRO which, whilst having a potential role in tumour suppression, may also be associated with AD.

Determination of Proteins with which USP25 interacts

Functional analysis of USP25 was performed with the aim of detecting the cellular proteins which interact with the USP25 protein through protein-protein interaction, using Yeast-Two-Hybrid (Y2H) approach. Saccharomyces

Cerevisiae yeast has well characterised ubiquitin activating, conjugating and ligating ezymatic machinery, capable of ubiquitinating human proteins (Scheffner et al. 1998). A cDNA library from human brain cloned in "prey" vector, was co-transfected to yeast cells with USP25 cloned in "bait" vector. Since ubiquitin cleaving activity of USP25 was proven (Groet et al. 2000), this technique has a theoretical chance of detecting the natural cellular substrates for ubiquitin cleavage and de-ubiquitination by USP25. Since the action of ubiquitin cleavage is very rapid (Wilkinson and Hochstrasser 1998), the cleavage and dissociation from its natural substrates for fully active USP25 could preclude the ability to detect the interaction through Y2H. In addition, the artificial cross-de-ubiquitination of yeast's own proteins by an overexpressed USP25 could theoretically be harmful for the yeast cell and/or for the molecular interactions required for the Y2H. For these reasons we performed Y2H using USP25-C178A, a site directed mutant we recently engineered (the mutation being of the key Cys residue in the Cys region) which abolishes the capacity for cleavage of ubiquitin by USP25, but should not interfere with the binding of USP25 to its natural ubiquitinated substrates, since this residue is conserved between all UCH-s and USP-s so far identified.

Y2H experiment using USP25-C178A cloned in the yeast two hybrid "bait" vector pAS2 (Clontech) co-transformed into yeast cells together with a human adult brain cDNA library in the "prey" vector pACT2 (Clontech). Interacting events were visualized by the activation of transcription of all three reporter genes: Ade2, Mel1 and His3. Interacting "prey" sequences were verified by PCR-sequencing on the ABI310 automated sequencer, using universal vector primers, and analysed by BLAST search on non-redundant genome and transcriptome sequence databases. The accession numbers of the sequences found to be interacting, from the GenBank database are given in the table.

Table 1. Summary of frequency and identities of specific interacting proteins from human brain with USP25-C178A, detected using Yeast-Two-Hybrid technique

Summary of Results by decreasing frequency of detection of bait	Number of specific independent clones "fished" by Y2H	Accession number
HHR23A	8 clones	D21235
SUMO-3	8 clones	NM 006937
human UBC9	5 clones	U66867
polyUbiquitin	4 clones	AB009010
Ubiquitin	3 clones	X04803
Ran BP2 protein	1 clone	NM 006267
Various ubiquitin-like precursors (1 or 2 clones each):	4 clones	
Other proteins (1 or 2 clones each)	10 clones	

hRAD23A (which is a homologue of yeast RAD23 protein) has been isolated as a primary interacting protein by the same Y2H technique using E6AP as a "bait" (Kumar et al. 1999). E6AP (Human Papilloma Virus E6 associated protein) functions as one of the two so far detected ubiquitin ligases (attaching ubiquitin and labelling for degradation) for the master tumour suppressor p53 (Scheffner et al. 1993). The p53 and HHR23A are the only two so far proven targets for this ubiquitin ligase (Kumar et al. 1999).. Since USP25 shows high rate of target preference for HHR23A (see Table 1), and both HHR23A and p53 are ubiquitinated by E6AP, it could mean that they are both de-ubiquitinated by USP25. The fact that Y2H with USP25 did not pick up p53 is understandable, because p53 is expressed in small traces (very low level) in normal tissues, and gets only accumulated and activated following DNA damage or other stimuli for programmed cell death (apoptosis). It is possible that the effect of overexpression of USP25 (and

other USP-s) may be neuronal toxicity and progressive neuronal cell death seen in AD.

Figure Legends

Figure 1. Identification of the Shared Region of Overlap (S.R.O.) for hemizygous deletions in 21q11-q21 in NSCLC.A: Cytogenetic map, *Not* I long range physical map (Ishikawa, *et al.*, 1993), YAC contig (Nizetic, *et al* 1994, Shinizu *et al* 1995 and Bosch *et al* 1996), and bacterial contig, (Groet *et al* 1998) are shown in consecutive horizontal layers, respectively, above the line showing the markers used in the LOH analysis (oval symbols). Markers are named as in Genome DataBase (prefix "D21" omitted). In the column under each marker an "X" (symbolizing LOH), "+" (an absence of LOH) and "U" (un-informative, homozygous result) for that marker in the set of eleven Croatian Tumour Bank (CTB) tumours, or in individual tumours #47 and #61 are shown. NT=not tested. For comparison with our data, markers used as probes on the genomic Southern blot of the NSCLC cell line, and/or in LOH analysis of fresh tumours in the study by Kohno and co-workers (Kohno *et al* 1998) are indicated above the empty bar symbolizing the homozygous deletion they found. In our data, hatched bars indicate hemizygous deletions, and black filled bars indicate segments showing absence of LOH or deletions. Squared symbols "X" and "+" stand for predominantly single and predominantly double signal, respectively, detected by FISH on interphase nuclei of the paraffin embedded sections of the tumour #61, when PAC clones named and indicated as bold lines in the PAC contig above the markers line, were used as probes.

Figure 2. Trapped exons (hatched rectangles) and exons deduced from overlapping sequence analysis (white rectangles) defining the exon-intron structure of the new gene USP25. Top half shows two PACs 73M5 and 135E14, also used as FISH probes in Fig. 1, which were the source of genomic DNA for exon trapping. Exon locations on the PACs are shown with vertical bars, and the 50 kbp scale bar refers to this part. Bottom half consists of overlapping cDNA fragments corresponding to exons above them,

drawn in the same scale, (500 bp scale bar is shown). Names of cDNA clones are as in dB-EST and UniGene databases, 824710 is the address of the clone in the IMAGE Consortium collection. The complete cDNA sequence for the whole gene is the new GenBank entry with the accession number AF 134213.

Figure 3. Comparison of protein sequences of USP25 to other eukaryotic members of the superfamily of USP-s. The protein BAP-1 is actually from the family of Ubiquitin C-terminal Hydrolases, a distinct sub-family of this superfamily, showing homology only in the single key aminoacids in the Cys and His domains. Two reports show the localisations of the highly homologous sequences for the HAUSP gene to 3p21 (Kashuba, *et al* 1997) and 16p13 (Robinson, *et al* 1998), respectively.

Figure 4. Demonstration of the de-ubiquitinating activity of USP25 on a model ubiquitin fusion protein. Western blot of an SDS-PAGE was detected using an anti-β-galactosidase antiserum. Lanes 1,2: the *E. coli* XL-1 blue cells alone (in all cases second line of the pair is +1PTG). Lanes 3,4: same cells co-transfected with the model fusion protein encoding plasmid pACYC-UB-Met-β-galactosidase protein, band labelled with an asterix). Lanes 5,6: as lanes 3,4 except pQE30-USP25 (full length USP25 gene cloned in the pQE30 expression vector) was added instead of pQE30. Lanes 7,8: same as lanes 3,4 except pRB105 (yeast de-ubiquitinating enzyme UBP2) was transfected instead of pQE30. Lanes 9,10: over-exposure of lanes 7,8. Note the presence of the 8kDa shorter, de-ubiquitinated Met-β-galactosidase (band labelled with a triangle).

References

Alzheimer, A. Allg.Z.Psychiat. 64:146-148(1907)

Amerik, A. Yu., Swaminathan, S., Krantz, B.A., Wilkinson, K.D., and Hochstrasser, M. (1997). *In vivo* disassembly of free polyubiquitin chains by yeast Ubp14 modulates rates of protein degradation by the proteasome. EMBO J. 16: 4826-4838.

Baker, R.T., Wang, X-W., Woollatt, E., White, J.A. and Sutherlands, G.R. Identification, functional characterisation, and chromosomal localisation of USP15, a novel Human USP related to Unp Oncoprotein, and a systematic nomenclature for hUSP's. Genomics 59, 264-274 (1999)

Beyreuther, K. and Masters, C.L. Tangle disentanglement. Nature, 383, 10 October 1996.

Borchelt, D. R. Metabolism of Presenilin 1: Influence of Presenilin 1 and Amyloid Precursor Prtotein Processing. Neurobiology of Aging. Vol. 19. No. 15. P 515-518, 1998.

Bosnar MH, Pavelic K, Krizanac S, Slobodnjak Z, Pavelic J. 1997. Squamous cell lung carcinomas: the role of nm23-H1 gene. J. Mol. Med. 75:609-613.

Brodsky, G., Otterson, G.A., Parry, B.B., Hart, I., Patterson, D., and Kaye, F.J. Localization of STCH to Human Chromosome 21q11.1. Genomics, 30:627-628, 1995.

Calhoun M.E. *et al.* Neuron loss in APP transgenic mice. Nature 395:755-756 (1998).

Cavailles, V., Dauvois, S., L'Horset, F., Lopez, G., Hoare, S., Kushner, P.J., and Parker, M.G. Nuclear factor RIP140 nodulates transcriptional activation by the estrogen receptor. EMBO J., 14:3741-3751, 1995.

Cummings C.J. et al., Neuron 24:879-892 (1999)

Everett, R.D., Meredith, M., Orr, A., Cross, A., Kathoria, M., and Parkinson, J.A novel ubiquitin-specific protease is dynamically associated with the PML nuclear domain and binds to a herpes virus regulatory protein. EMBO J., 16:1519-1530, 1997

Fero, M.L., Randel, E., Gurley, K.E., Roberts, J.M., and Kemp, C.J. The murine gene p27 ^{Kio1} is haplo-insufficient for tumour suppression. Nature, 396:177-180, 1998.

Fraser *et al.* Presenilin 1 is Actively Degraded by the 265 Proteasome. Neurobiology of Aging. Vol. 19. No. 15. P 519-521, 1998.

Fujiwara, T., Saito, A.., Suzuki, M., Shinomiya, H., Suzuki, T., Takahashi, E., Tanigami, A., Ichiyama, A., Chung, C.H., Nakamura, Y., and Tanaka, K. Identification and chromosomal assignment of USP1, a novel gene encoding a human ubiquitin-specific protease. Genomics, 54:155-158, 1998.

Games, D. *et al.* Alzheimer-type neuropathology in transgenic mice overexpressing V717F beta-amyloid precursor protein. Nature 373:523-527 (1995).

Gardiner K. 1996. Base composition and gene distribution: critical patterns in mammalian genome organisation. Trends Genet. 12:519-524.

Goate, A.M *et al.* Predisposing Locus for Alzheimer's Disease on Chromosome 21. The Lancet, p 352-355, 18 February 1989.

Goedert M., Wischik C.M., Crowther R.A., Walker J.E. and Klug A. Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau. Proc. Natl. Acad. Sci. USA 85:4051-4055 (1988).

Gong L. et al. J.Biol.Chem. 272(45):28198-28201 (1997)

Gong L. et al. J.Biol.Chem. 275(12):14212-14216 (2000b)

Groet, J., Ives, J.H., South A.P., Baptista, P.B., Jones, T.A., Yaspo, M-L., Lehrach, H., Potier, M-C., Van Broeckhoven, C., Nizetic, D. Bacterial contig map of the 21q11 region associated with Alzheimer's disease and abnormal myelopoiesis in Down syndrome. Genome Res., 8:385-398, 1998.

Groet, J. Ives, J.H., Jones, T.A., Danton, M., Flomen, R.H., Sheer, D., Hrasvcan, R., Pavelic, K., and Nizetic, D. Narrowing of the Region of Allelic Loss in 21q 11-21 in squamous Non-small cell Lung Carcinoma and Cloning of a Novel Ubiquitin-specific Protease Gene from the Deleted Segment Genes Chromosomes Cancer 27:153-161 (2000).

Hansen-Hagge TE, Janssen JWG, Hameister H, Papa FR, Zechner U, Seriu T, Jauch A, Becke D, Hochstrasser M, Bartram CR. 1998. An evolutionarily conserved gene on human chromosome 5q33-q34, UBH1, encodes a novel deubiquitinating enzyme. Genomics 49:411-418.

Ichikawa, H., Hosoda, F., Arai, Y., Shimizu, K., Ohira, M., and Ohki, M. A *Not* I restriction map of the entire long arm of human chromosome 21. Nat. Genet., 4:361-366 1993.

Jensen DE, Proctor M, Marquis ST, Gardner HP, Ha SI, Chodosh LA, Ishov AM, Tommerup N, Vissing H, Sekido Y, Minna J, Borodovsky A, Schultz DC, Wilkinson KD, Maul GG, Barlev N, Berger SL, Predergast GC, Rauscher FJ III. 1998. BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1 - mediated cell growth suppression. Oncogene 16:1097-1112.

Kamitani T. et al. J.Biol.Chem. 272(45):28557-28562 (19

Kamitani T. et al. J.Biol.Chem. 273(18):11349-11353 (1998a)

Kamitani T. et al. J.Biol.Chem. 273(6):3117-3120 (1998b)

Kashuba, V.I., Grizatullin, R.Z, Protopopov AI, Allikmets R, Korolev S., Li J, Boldog R, Tory K, Zabarovska V, Marcsek Z, Sumegi J, Klein G, Zabarovsky ER, Kisselev L. 1997. NotI linking/jumping clones of human chromosome 3: mapping of the TFRC, RAB7 and HAUSP genes to regions rearranged in leukemia and deleted in solid tumours. FEBS Lett 419:181-185.

Kohno, T., Kawanishi, M., Matsuda, S., Ichikawa, H., takada, M., Ohki, M., Yammamoto, T., and Yokota. J. Homozygous deletion and frequent allelic loss of the 21q11.1-q21.1 region including the ANA gene in human lung carcinoma. Genes Chromosom. Cancer, 21:236-243, 1998.

Kubuttat M.H.G. et al. Nature 387:299-303 (1997)

Kumar S. et al. J.Biol.Chem. 274(26)18785-18792 (1999)

Lane D. 1998. Awakening angels. Nature 394:616-617.

Lane N.J., Balbo A. and Rapoport S.I. A fine structural study of the hippocampus and dorsal root ganglion in mouse trisomy 16, a model for Down's Syndrome. Cell Biology International 20(10):673-680 (1996).

Lapenta V. et al. Genomics 40:362-366 (1997)

Lee G.W. et al. J.Biol.Chem. 273(11):6503-6507 (1998)

Leroy, *E. et al.* The ubiquitin pathway in Parkinson's Disease. Nature, 395, 1 October, 1998.

Lowe, J. *et al.* Ubiquitin Carboxyl-Terminal Hydrolase (PGP 9.5) is selectively present in ubiquitinated Inclusion bodies characteristic of human neurodegenerative diseases. Journal of Pathology, 161: 153-160, 1990.

Mahajan R. et al. Cell 88:97-107 (1997)

Manetto, V. *et al.* Ubiquitin is associated with abnormal cytoplastic filaments characteristic of neurodegenerative diseases. Proc. Natl. Acad. Sci USA 85 p4501-4505, 1988.

Mannen H. et al. Biochem.Biophys.Res.Comm. 222:178-180 (1996)

Masutani C. et al. EMBO J. 13(8):1831-1843 (1994)

Matunis M.J. et al. J.Cell Biol. 140(3):499-509 (1998)

Meluh P.B. and Koshland D., Mol. Biol. Cell 6:793-807 (1995)

Mori, H. *et al.* Ubiquitin Is a Component of Paired Helical Filaments in Alzheimer's Disease. Science, 235, p1641-1644, 27 March 1987.

Morishima-Kawashima, M. *et al.* Ubiquitin is conjugated with amino-terminally processed Tau in paired helical filaments. Neuron, 10:1151-1160 (1993).

Nizetic, D., Gellen L., Hamvas R., Mott R., Grigoriev A., Vatcheva R., Zehetner G., Yaspo M.L., Dutriaux A., Lopes C., Delabar J.M., Van Broeckhoven C., Potier M.C. & Lehrach H.; An integrated YAC overlap and "cosmid-pocket" map of the human chromosome 21. Human Molecular Genetics 3(5):759-770 (1994).

Okura T. et al. J.Immunol. 157:4277-4281 (1996)

Perry, G. *et al.* Ubiquitin is detected in neurofibrillary tangles and senile plaque neurites of Alzheimer's disease brains. Proc. Natl. Acad. Sci. USA. 84, p3033-3036, 1987.

Reeves, R.H. *et al.* A mouse model for Down syndrome exhibits learning and behaviour deficits. Nature Genetics, 11:177-184 (1995).

Richards, S-J *et al.* Transplants of mouse trisomy 16 hippocampus provide a model of Alzheimer's disease neuropathology. The EMBO Journal 10:297-303, (1991).

Robinson PA, Lomonte P, Leek P, Markham AF, Everett RD, 1998. Assignment of herpesvirus-associated ubiquitin-specific protease gene HAUSP to human chromosome band 16p 13.3 by in situ hybridisation, Cytogenet Cell Genet 83:100.

Saitoh H. and Hinchey J., J.Biol.Chem. 275(9)6252-6258 (2000)

Sambrook, J., Fritsch, E.F., & Mamiatis, T Molecular Cloning:a laboratory manual 2nd ed. Coldspring Harbor, N.Y. 1989.

Saudou F. et al. Cell 95:55-66 (1998)

Scheffner M. et al. Cell 75:495-505 (1993)

Scheffner M. et al. The Ubiquitin-Conjugation System. In: J.M. Peters, D. Finley and R. Harris (eds.)

Schwarz S.E. et al. Proc.Natl.Acad.Sci.USA 95:560-564 (1998)

Shen Z. et al. Genomics 36:271-279 (1996a)

Shen Z. et al. Genomics 36:183-186 (1996b)

Shimizu, N., Antonarakis, S., VanBroeckhoven, C., Patterson, D., Gardiner, K., Nizetic, D., Creau, N., Delabar, J., Korenberg, J., Reeves, R., Doering, J., Ritter, O., and Cuticchia, J. Report of the fifth international workshop on human chromosome 21 mapping 1994. Cytogenet. Cell Genet., 70:147-182, 1995.

Shoji M. *et al.* Accumulation of amyloid beta protein in transgenic mice. Neurobiology of aging Vol 19(15):S59-S63 (1998).

Spaventi R., Pecur, L., Pavelic, K., Pavelic, Z.P., Spaventi, S. and Sambrook, P.J. Human Tumour Bank in Croatia: as possible model for a small bank as part of the future European Tumour Bank Network. Eur. J. Cancer 30A, 419-1994.

St George-Hyslop, P.H. *et al.* The Genetic Defect Causing Familial Alzheimer's Disease Maps on chromosome 21. Science, 235, p885-890, 20 February 1987.

Swanson DA, Freund CL, Ploder L, McInnes RR, Valle D. 1996. A ubiquitin C-terminal hydrolase gene on the proximal short arm of the X chromosome: implications for X-linked retinal disorders. Hum Mol Genet 5:533-538.

Teller J.K., Russo C., DeBusk L.M., Angelini G., Zaccheo D., Dagna-Bricarelli F., Scartezzini P., Bertolini S., Mann D.M.A., Tabaton M. and Gambetti P. Presence of soluble amyloid beta-peptide precedes amyloid plaque formation in Down's Syndrome. Nature Medicine, 2(1):93-95 (1996).

Terry RD & Katzmann R, Ann. Neurol. 114:497-506(1983)

The *C. elegans* Sequencing Consortium, 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282:2012-2018.

Valero, R., Martany, G., Gonzalez-Angulo, O., Gonzalez-Gonzalez, G., Puelles, L and Gonzales-Duarte, R. USP25, a Novel Gene Encoding a Deubiquitinating Enzyme, is located in the Gene-Poor Region 21q11.2. Genomics 62: 395-405 (1999).

Wilkinson, K.D. (1997). Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. *FASEB J.* 11: 1245-1256.

Wilkinson, K.D., and Hochstrasser, M. Deubiquitinating enzymes. In: J.M. Peters, D. Finley and R. Harris (eds.), Ubiquitin and the Biology of the Cell, Plenum Press, New York, 1998.

Wischik, C. *et al.* Isolation of a fragment of Tau derived from the core of the paired helical filament of Alzheimer disease. Proc. Natl. Acad. Sci. USA 85:4506-4510 (1998).

Wolozin, B. *et al* Regulation of Apoptosis by Presenilin 1. Neurobiology of Aging. Vol. 19 (15) p 523-527, 1998.

Yaspo M-L, Gellen L, Mott R, Korn B, Nizetic D, Poustka A, Lehrach H. 1995. Model for a transcript map of human chromosome 21: isolation of new coding sequences from exon and enriched cDNA libraries. Hum. Mol. Genet 4:1291-1304.

CLAIMS

1. Use of ubiquitin specific protease (USP), an inhibitor thereof or the gene therefor, or a specific inhibitor or promoter of such a gene, in the manufacture of a composition for use in the treatment, diagnosis or prophylaxis of AD.

2. Use of ubiquitin-like specific protease an inhibitor thereof or the gene therefor, or a specific inhibitor or promoter of such a gene, in the manufacture of a composition for use in the treatment, diagnosis or prophylaxis of AD.

3. Use according to claim 1 or claim 2 in which the protease is a protein having ubiquitin specific protease or ubiquitin-like specific protease activity which includes the following three domains having the specified sequences of the general formulae I, II and III:

$$GXX^1\ NX^2GX^3TCX^4X^5X^6X^7X^8X^9QX^{10}X^{11} \quad \text{I}$$

$$QX^{12}DX^{13}X^{14}EX^{15}X^{16}X^{17}X^{18}X^{19}X^{20}X^{21}X^{22}X^{23}X^{24}X^{25} \quad \text{II}$$

$$YX^{26}LX^{27}X^{28}X^{29}X^{30}X^{31}HX^{32}GX^{33}X^{34}X^{35}X^{36}X^{37}GHYX^{38}X^{39}X^{40} \quad \text{III}$$

in which

X is the residue of a non-polar amino acid

X^1 is the residue of an amino acid with an uncharged or basic R group

X^2 is the residue of an amino acid with an uncharged R group

X^3 is the residue of an amino acid with an uncharged R group

X^4 is the residue of an amino acid having a relatively large uncharged R group

X^5 is the residue of an amino acid having an uncharged R group

X^6 is the residue of an amino acid having a relatively polar uncharged R group

X^7 is the residue of an amino acid having an uncharged R group

X^8 is the residue of an amino acid having a non-polar R group

X^9 is the residue of an amino acid having an uncharged R group

X^{10} is the residue of an amino acid having an uncharged R group

X^{11} Is the residue of an amino acid having an uncharged R group

X^{12} is Q or H

X^{13} is the residue of an amino acid having a polar R group

X^{14} is the residue of an amino acid having an uncharged or basic R group

X^{15} is the residue of an amino acid having an uncharged R group

X^{16} is the residue of an amino acid having an uncharged R group

X^{17} is the residue of an amino acid having an uncharged polar R group or a basic R group

X^{18} is the residue of an amino acid having an uncharged or a basic R group

X^{19} is the residue of an amino acid having a polar R group

X^{20} is the residue of an amino acid having an uncharged R group

X^{21} is the residue of an amino acid having an uncharged polar R group or an acidic R group

X^{22} is the residue of an amino acid having an uncharged polar R group or a basic R group

X^{23} is the residue of an amino acid having a non-polar R group

X^{24} is the reside of an amino acid having an uncharged polar R group or an acidic R group

X^{25} is the residue of any amino acid

X^{26} is the residue of any amino acid

X^{27} is the residue of an uncharged or a basic amino acid

X^{28} is the residue of an amino acid having a non-polar R group

X^{29} is the residue of an amino acid having a non-polar R group

X^{30} is the residue of an amino acid having a non-polar R group

X^{31} is the residue of an amino acid having an uncharged R group

X^{32} is the residue of an amino acid having an uncharged polar R group or an acidic R group

X^{33} is the residue of an amino acid having an uncharged or an acidic R group

X^{34} is the residue of an amino acid having an uncharged R group

X^{35} is residue of an amino acid having an uncharged or a basic R group

X^{36} is the residue of an amino acid having an uncharged or a basic R group

X^{37} is the a bond or the residue of an amino acid having an uncharged polar R group or a basic R group, or is a bond

X^{38} is the residue of an amino acid having an uncharged R group

X^{39} is the residue of an amino acid having an uncharged R group and

X^{40} is the residue of an amino acid having an uncharged R group.

4. Use according to claim 3 in which:

X is F or L;

X^1 is K or R;

X^2 is selected from V, S, U, A, L and F;

X^3 is A or N;

X^4 is Y or W;

X^5 is selected from F, M, L and C;

X^6 is N or S;

X^7 is selected from A, C and S;

X^8 is selected from V, L and I;

X^9 is selected from V, L, I and S;

X^{10} is selected from S, V, Q, A and T;

X^{11} is Y or L;

X^{12} is Q;

X^{13} is V or A;

X^{15} is selected from F, L, M and V;

X^{19} is selected from L, I and V;

X^{20} is L;

X^{21} is D;

X^{22} is W;

X^{23} is selected from L, I and V;

X^{24} is selected from E, Q, D and A;

X^{25} is D;

X^{27} is selected from H, Y, V and F;

X^{28} is A or G;

X^{29} is selected from L, I and V;

X^{30} is selected from L, I and V;

X^{31} is selected from L, I, V and M;

X^{32} is selected from E, S and C;

X^{33} is selected from D, E, Q and S;

X^{34} is selected from L, I, V, A and G;

X^{35} is selected from N, H, S and P;

X^{37} is a bond;

X^{38} is selected from Y, W, T, I and V;

X^{39} is selected from A, S, T and V; and

X^{40} is selected from Y, L, S and I.

5. Use according to claim 3 or claim 4 in which:

X^{14} is selected from S, T, Q, L and H;

X^{16} is selected from T, L, N, F and C;

X^{17} is selected from H, T, R, N and Q;

X^{18} is selected from K, L, I, V, S, C and Y;

X^{26} is selected from R, E, V, N, G, D and I; and

X^{36} is selected from A, R, N, T, G and S.

6. Use according to claim 3 in which the three domains have the sequences I.D. Nos 2-4, respectively.

7. Use according to any preceding claim in which the protease has sequence ID 1, or is a fragment thereof with ubiquitin specific protease or ubiquitin-like specific protease activity, or a homologue from a non-human animal.

8. Use according to any of claims 1 to 6 in which the protease has sequence ID 5.

9. Use according to any of claims 1 to 6 in which the gene is used and the gene comprises sequence ID No. 6.

10. Use according to any of claims 1 to 6 in which the gene is used and the gene comprises sequence ID No 7.

11. An *in vitro* method of diagnosis of AD in a human or animal in which a ubiquitin specific protease, or the gene therefor is detected in a sample taken

from a patient suspected of suffering from AD, wherein an alteration in the level of expression of protease or in the sequence of the translated protease as compared to normal levels is detected.

12. An *in vitro* method of diagnosis of AD in a human or animal in which a ubiquitin-like specific protease, or the gene therefor is detected in a sample taken from a patient suspected of suffering from AD, wherein an alteration in the level of expression of protease or in the sequence of the translated protease as compared to normal levels is detected

13. A method for investigating the pathogenesis of AD by the use of a USP, or the gene therefore.

14. A method for investigating the pathogenesis of AD by the use of a ubiquitin-like specific protease.

15. A method according to any of claims 11 to 14 in which the protease is as defined in any of claims 3 to 8.

16. A method of synthesizing a protease having Cys, QQD and His domains specified in sequence ID's Nos. 2, 3 and 4, respectively in which nucleic acid encoding the protease is introduced into a microorganism or a cell-line in a form in which it can be replicated, transcribed and translated, and the microorganism or cell-line is cultured under conditions whereby the nucleic acid is replicated, transcribed and translated to form the protease, and the protease is collected.

17. A method according to claim 16 in which the nucleic acid introduced into the microorganism, or cell-line, as the case may be, includes the sequence represented by ID. No. 6 or No. 7.

18. A nucleic acid construct comprising a USP gene encoding a protein product having Cys, QQD and His regions as specified in sequence ID's Nos. 2, 3 and 4 and having an origin of replication.

19. A construct according to claim 18 formed from a vector having an origin of replication and preferably also a promoter for initiation of translation.

20. A construct according to claim 18 or 19 in which the gene has sequence ID. No. 6 or No. 7.

21. A nucleic acid construct encoding a product having a contiguous sequence of at least 10 residues having at least 50% identity with a contiguous sequence of residues of sequence ID 1, and an origin of replication.

22. A construct according to claim 21 in which the product has a contiguous sequence of at least 20, for instance 50 or more, residues having at least 20%, preferably at least 50%, identity with a contiguous sequence of residues of sequence ID 1.

23. A construct according to claim 21 or 22 in which the product has two or more, for instance three or four, contiguous sequences of at least 10 bases having at least 5% identity with respective contiguous sequences of sequence ID 1.

24. A transformed microorganism containing the construct of any of claims 18 to 23.

25. A transfected cell-line containing the construct of any of claims 18 to 23.

26. A cell-line according to claim 25 which is derived from neuroblasts, neuroepithelial cells or neurones.

27. USP produced by a method according to claim 16 or 17.

28. A protein product which is not a functional protease which has a contiguous sequence of at least 10 residues having at least 50% identity with a contiguous sequence of sequence ID 1.

29. A protein product according to claim 28 which has a contiguous sequence of at least 20 residues, for instance 50 or more, having at least 20%, preferably at least 50%, identity with a contiguous sequence of sequence ID 1.

30. A protein product according to claim 28 or 29 which has contiguous sequences which differ from those of sequence ID's 2, 3 and/or 4 in respect of one, two or three residues, preferably including the Cys residue which is the 9th residue of sequence ID 2, and/or one or more of the QQD residues which are the 1st to 3rd residues of sequence ID 3, and/or the His residue which is

the 17th residue of sequence ID 3, most preferably the Cys residue of sequence ID 3.

31. A product according to any of claims 28 to 30 which is a competitive inhibitor of ubiquitin specific protease or ubiquitin-like specific protease activity.

32. A protein having sequence ID 1 in substantially isolated form.

33. A substantially isolated protein which is characterised by its ability to interact with one or more and preferably all those of the proteins having the sequences of GenBank Accession Nos. D21235 (HHR23A), NM 006937 (SUMO-3), U 66867 (UBC9), AB 009010 (poly Ubiquitin) and X 04803 (Ubiquitin), as determined using the encoding gene in a yeast-two-hybrid method.

34. A non-human animal model useful to investigate the mechanism of AD involving USP, having the gene for USP homologous to that found at human chromosome 21q11-21 knocked out, hemi- or homozygously, is polysomic for that gene, or has that gene modified by disrupting the transcriptional or translational control whereby the animal has reduced or increased levels of the corresponding USP.

35. A non-human animal according to claim 34 which is a mouse.

Figure 1

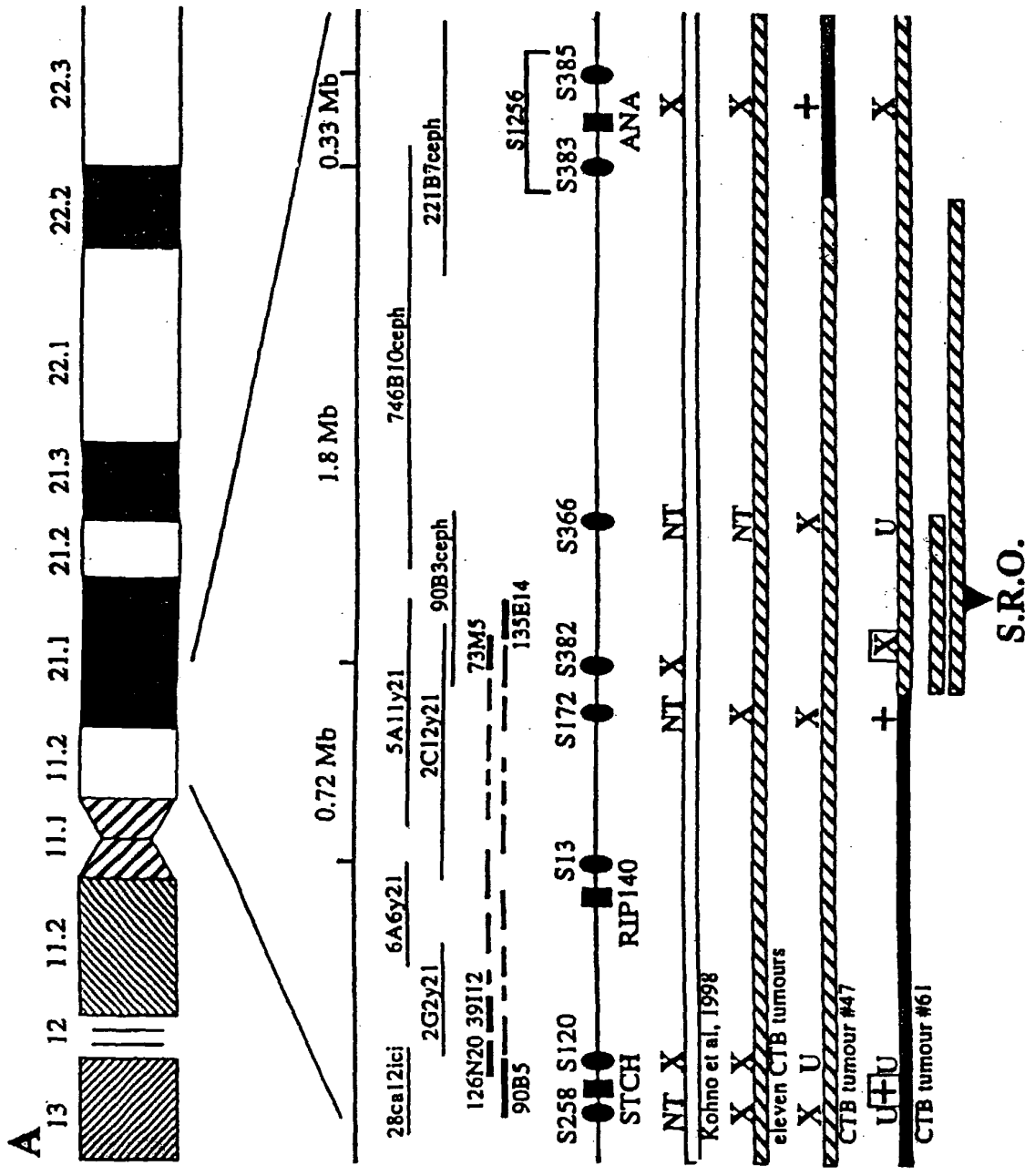

A
13
12
11.2
11.1
11.2
21.1
21.2
21.3
22.1
22.2
22.3
0.72 Mb
1.8 Mb
0.33 Mb
28ca12ici
6A6y21
5A11y21
746B10ceph
2G2y21
2C12y21
90B3ceph
221B7ceph
126N20 39I12
73M5
S1256
90B5
135E14
S383
S385
ANA
S258 S120
S13
S172 S382
S366
STCH
RIP140
NT X
NT X
NT
X
Kohno et al, 1998
eleven CTB tumours
CTB tumour #47
CTB tumour #61
S.R.O.

Figure 2

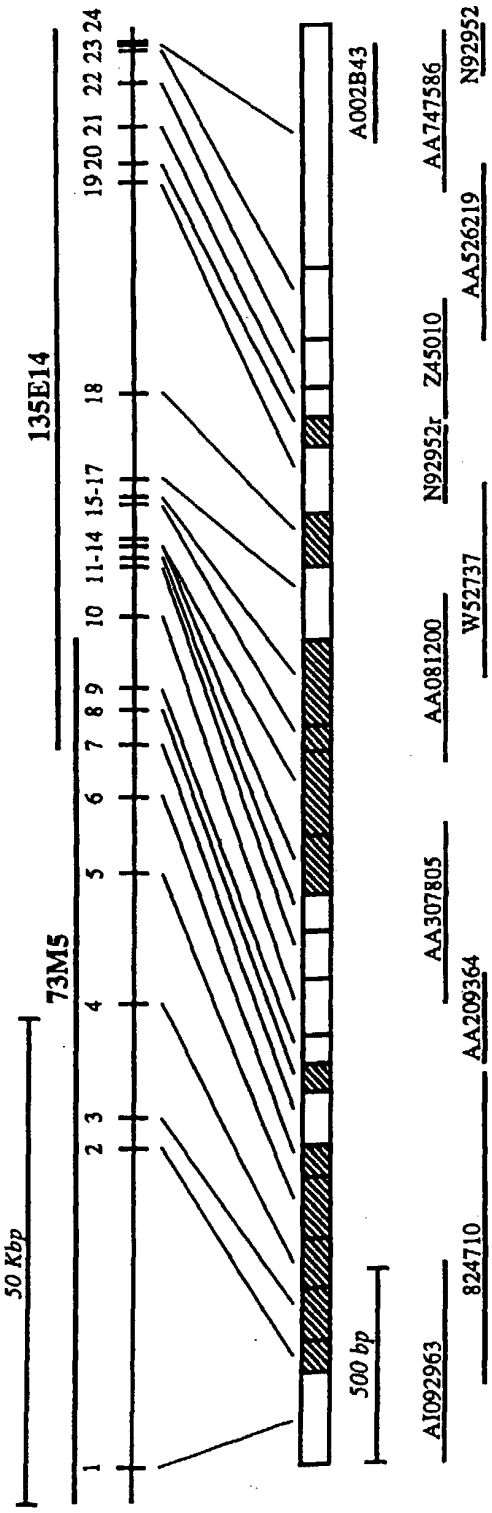

50 Kbp
73M5
135E14
1
2 3
4
5
6
7 8 9
10
11-14
15-17
18
19 20 21
22 23 24
500 bp
AI092963
824710
AA209364
AA307805
AA081200
W52737
N92952r
Z45010
AA526219
AA747586
N92952
A002B43

Figure 3

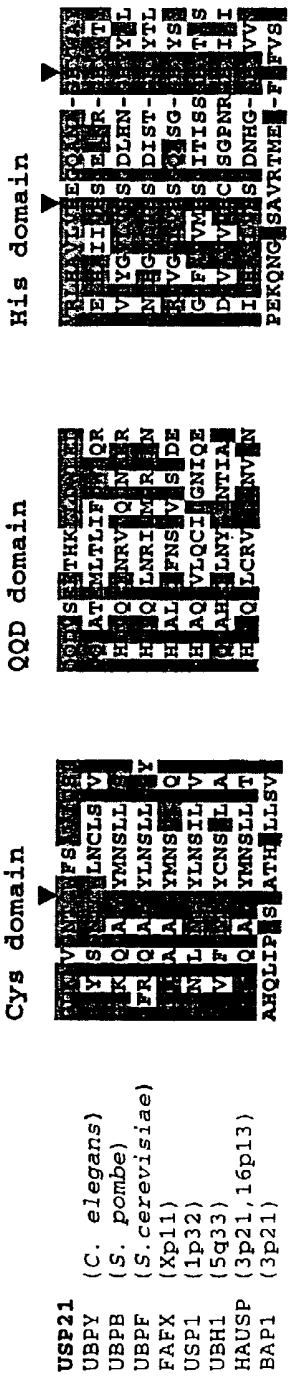
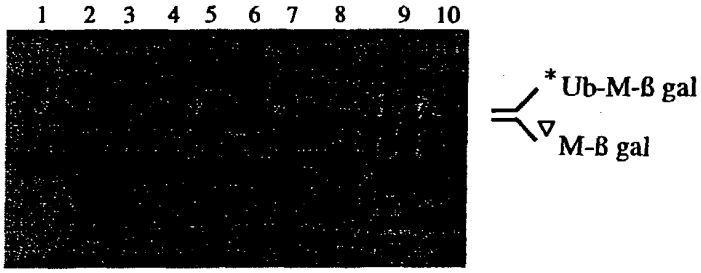

Cys domain
QQD domain
His domain
USP21
UBPY (C. elegans)
UBPB (S. pombe)
UBPF (S.cerevisiae)
FAFX (Xp11)
USP1 (1p32)
UBH1 (5q33)
HAUSP (3p21,16p13)
BAP1 (3p21)

Figure 4



SEQUENCE LISTING

<110> School of Pharmacy

<120> Diagnosis and Treatment of Alzheimer's Disease

<130> HMJ03220WO

<140> PCT/GB 0002423
<141> 2000-06-22

<150> 9914589.8
<151> 1999-06-22

<150> 0008162.0
<151> 2000-04-03

<160> 7

<170> PatentIn Ver. 2.1

<210> 1
<211> 1055
<212> PRT
<213> Homo sapiens

<400> 1

```
Met Thr Val Glu Gln Asn Val Leu Gln Gln Ser Ala Ala Gln Lys His
 1               5                  10                  15

Gln Gln Thr Phe Leu Asn Gln Leu Arg Glu Ile Thr Gly Ile Asn Asp
            20                  25                  30

Thr Gln Ile Leu Gln Gln Ala Leu Lys Asp Ser Asn Gly Asn Leu Glu
        35                  40                  45

Leu Ala Val Ala Phe Leu Thr Ala Lys Asn Ala Lys Thr Pro Gln Gln
    50                  55                  60

Glu Glu Thr Thr Tyr Tyr Gln Thr Ala Leu Pro Gly Asn Asp Arg Tyr
65                  70                  75                  80

Ile Ser Val Gly Ser Gln Ala Asp Thr Asn Val Ile Asp Leu Thr Gly
                85                  90                  95

Asp Asp Lys Asp Asp Leu Gln Arg Ala Ile Ala Leu Ser Leu Ala Glu
            100                 105                 110
```

```
Ser Asn Arg Ala Phe Arg Glu Thr Gly Ile Thr Asp Glu Glu Gln Ala
        115                 120                 125

Ile Ser Arg Val Leu Glu Ala Ser Ile Ala Glu Asn Lys Ala Cys Leu
    130                 135                 140

Lys Arg Thr Pro Thr Glu Val Trp Arg Asp Ser Arg Asn Pro Tyr Asp
145                 150                 155                 160

Arg Lys Arg Gln Asp Lys Ala Pro Val Gly Leu Lys Asn Val Gly Asn
                165                 170                 175

Thr Cys Trp Phe Ser Ala Val Ile Gln Ser Leu Phe Asn Leu Leu Glu
            180                 185                 190

Phe Arg Arg Leu Val Leu Asn Tyr Lys Pro Pro Ser Asn Ala Gln Asp
        195                 200                 205

Leu Pro Arg Asn Gln Lys Glu His Arg Asn Leu Pro Phe Met Arg Glu
    210                 215                 220

Leu Arg Tyr Leu Phe Ala Leu Leu Val Gly Thr Lys Arg Lys Tyr Val
225                 230                 235                 240

Asp Pro Ser Arg Ala Val Glu Ile Leu Lys Asp Ala Phe Lys Ser Asn
                245                 250                 255

Asp Ser Gln Gln Gln Asp Val Ser Glu Phe Thr His Lys Leu Leu Asp
            260                 265                 270

Trp Leu Glu Asp Ala Phe Gln Met Lys Ala Glu Glu Glu Thr Asp Glu
        275                 280                 285

Glu Lys Pro Lys Asn Pro Met Val Glu Leu Phe Tyr Gly Arg Phe Leu
    290                 295                 300

Ala Val Gly Val Leu Glu Gly Lys Lys Phe Glu Asn Thr Glu Met Phe
305                 310                 315                 320

Gly Gln Tyr Pro Leu Gln Val Asn Gly Phe Lys Asp Leu His Glu Cys
                325                 330                 335

Leu Glu Ala Ala Met Ile Glu Gly Glu Ile Glu Ser Leu His Ser Glu
            340                 345                 350

Asn Ser Gly Lys Ser Gly Gln Glu His Trp Phe Thr Glu Leu Pro Pro
        355                 360                 365
```

```
Val Leu Thr Phe Glu Leu Ser Arg Phe Glu Phe Asn Gln Ala Leu Gly
    370                 375                 380

Arg Pro Glu Lys Ile His Asn Lys Leu Glu Phe Pro Gln Val Leu Tyr
385                 390                 395                 400

Leu Asp Arg Tyr Met His Arg Asn Arg Glu Ile Thr Arg Ile Lys Arg
                405                 410                 415

Glu Glu Ile Lys Arg Leu Lys Asp Tyr Leu Thr Val Leu Gln Gln Arg
            420                 425                 430

Leu Glu Arg Tyr Leu Ser Tyr Gly Ser Gly Pro Lys Arg Phe Pro Leu
        435                 440                 445

Val Asp Val Leu Gln Tyr Ala Leu Glu Phe Ala Ser Ser Lys Pro Val
    450                 455                 460

Cys Thr Ser Pro Val Asp Asp Ile Asp Ala Ser Ser Pro Pro Ser Gly
465                 470                 475                 480

Ser Ile Pro Ser Gln Thr Leu Pro Ser Thr Thr Glu Gln Gln Gly Ala
                485                 490                 495

Leu Ser Ser Glu Leu Pro Ser Thr Ser Pro Ser Ser Val Ala Ala Ile
            500                 505                 510

Ser Ser Arg Ser Val Ile His Lys Pro Phe Thr Gln Ser Arg Ile Pro
        515                 520                 525

Pro Asp Leu Pro Met His Pro Ala Pro Arg His Ile Thr Glu Glu Glu
    530                 535                 540

Leu Ser Val Leu Glu Ser Cys Leu His Arg Trp Arg Thr Glu Ile Glu
545                 550                 555                 560

Asn Asp Thr Arg Asp Leu Gln Glu Ser Ile Ser Arg Ile His Arg Thr
                565                 570                 575

Ile Glu Leu Met Tyr Ser Asp Lys Ser Met Ile Gln Val Pro Tyr Arg
            580                 585                 590

Leu His Ala Val Leu Val His Glu Gly Gln Ala Asn Ala Gly His Tyr
        595                 600                 605

Trp Ala Tyr Ile Phe Asp His Arg Glu Ser Arg Trp Met Lys Tyr Asn
    610                 615                 620
```

```
Asp Ile Ala Val Thr Lys Ser Ser Trp Glu Glu Leu Val Arg Asp Ser
625                 630                 635                 640

Phe Gly Gly Tyr Arg Asn Ala Ser Ala Tyr Cys Leu Met Tyr Ile Asn
                645                 650                 655

Asp Lys Ala Gln Phe Leu Ile Gln Glu Glu Phe Asn Lys Glu Thr Gly
            660                 665                 670

Gln Pro Leu Val Gly Ile Glu Thr Leu Pro Pro Asp Leu Arg Asp Phe
        675                 680                 685

Val Glu Glu Asp Asn Gln Arg Phe Glu Lys Glu Leu Glu Glu Trp Asp
    690                 695                 700

Ala Gln Leu Ala Gln Lys Ala Leu Gln Glu Lys Leu Leu Ala Ser Gln
705                 710                 715                 720

Lys Leu Arg Glu Ser Glu Thr Ser Val Thr Thr Ala Gln Ala Ala Gly
                725                 730                 735

Asp Pro Glu Tyr Leu Glu Gln Pro Ser Arg Ser Asp Phe Ser Lys His
            740                 745                 750

Leu Lys Glu Glu Thr Ile Gln Ile Ile Thr Lys Ala Ser His Glu His
        755                 760                 765

Glu Asp Lys Ser Pro Glu Thr Val Leu Gln Ser Ala Ile Lys Leu Glu
    770                 775                 780

Tyr Ala Arg Leu Val Lys Leu Ala Gln Glu Asp Thr Pro Pro Glu Thr
785                 790                 795                 800

Asp Tyr Arg Leu His His Val Val Val Tyr Phe Ile Gln Asn Gln Ala
                805                 810                 815

Pro Lys Lys Ile Ile Glu Lys Thr Leu Leu Glu Gln Phe Gly Asp Arg
            820                 825                 830

Asn Leu Ser Phe Asp Glu Arg Cys His Asn Ile Met Lys Val Ala Gln
        835                 840                 845

Ala Lys Leu Glu Met Ile Lys Pro Glu Glu Val Asn Leu Glu Glu Tyr
    850                 855                 860

Glu Glu Trp His Gln Asp Tyr Arg Lys Phe Arg Glu Thr Thr Met Tyr
865                 870                 875                 880
```

```
Leu Ile Ile Gly Leu Glu Asn Phe Gln Arg Glu Ser Tyr Ile Asp Ser
                885                 890                 895

Leu Leu Phe Leu Ile Cys Ala Tyr Gln Asn Asn Lys Glu Leu Leu Ser
            900                 905                 910

Lys Gly Leu Tyr Arg Gly His Asp Glu Glu Leu Ile Ser His Tyr Arg
        915                 920                 925

Arg Glu Cys Leu Leu Lys Leu Asn Glu Gln Ala Ala Glu Leu Phe Glu
    930                 935                 940

Ser Gly Glu Asp Arg Glu Val Asn Asn Gly Leu Ile Ile Met Asn Glu
945                 950                 955                 960

Phe Ile Val Pro Phe Leu Pro Leu Leu Leu Val Asp Glu Met Glu Glu
                965                 970                 975

Lys Asp Ile Leu Ala Val Glu Asp Met Arg Asn Arg Trp Cys Ser Tyr
            980                 985                 990

Leu Gly Gln Glu Met Glu Pro His Leu Gln Glu Lys Leu Thr Asp Phe
        995                1000                1005

Leu Pro Lys Leu Leu Asp Cys Ser Met Glu Ile Lys Ser Phe His Glu
   1010                1015                1020

Pro Pro Lys Leu Pro Ser Tyr Ser Thr His Glu Leu Cys Glu Arg Phe
1025                1030                1035                1040

Ala Arg Ile Met Leu Ser Leu Ser Arg Thr Pro Ala Asp Gly Arg
               1045                1050                1055
```

<210> 2
<211> 18
<212> PRT
<213> Homo sapiens

<400> 2

```
Gly Leu Lys Asn Val Gly Asn Thr Cys Trp Phe Ser Ala Val Ile Gln
  1               5                  10                  15

Ser Leu
```

<210> 3
<211> 17
<212> PRT
<213> Homo sapiens

<400> 3
Gln Gln Asp Val Ser Glu Phe Thr His Lys Leu Leu Asp Trp Leu Glu
1 5 10 15

Asp

<210> 4
<211> 21
<212> PRT
<213> Homo sapiens

<400> 4
Tyr Arg Leu His Ala Val Leu Val His Glu Gly Gln Ala Asn Ala Gly
1 5 10 15

His Tyr Trp Ala Tyr
20

<210> 5
<211> 1087
<212> PRT
<213> Homo sapiens

<400> 5
Met Thr Val Glu Gln Asn Val Leu Gln Gln Ser Ala Ala Gln Lys His
1 5 10 15

Gln Gln Thr Phe Leu Asn Gln Leu Arg Glu Ile Thr Gly Ile Asn Asp
20 25 30

Thr Gln Ile Leu Gln Gln Ala Leu Lys Asp Ser Asn Gly Asn Leu Glu
35 40 45

Leu Ala Val Ala Phe Leu Thr Ala Lys Asn Ala Lys Thr Pro Gln Gln
50 55 60

Glu Glu Thr Thr Tyr Tyr Gln Thr Ala Leu Pro Gly Asn Asp Arg Tyr

65 70 75 80

Ile Ser Val Gly Ser Gln Ala Asp Thr Asn Val Ile Asp Leu Thr Gly
85 90 95

Asp Asp Lys Asp Asp Leu Gln Arg Ala Ile Ala Leu Ser Leu Ala Glu
100 105 110

Ser Asn Arg Ala Phe Arg Glu Thr Gly Ile Thr Asp Glu Glu Gln Ala
115 120 125

Ile Ser Arg Val Leu Glu Ala Ser Ile Ala Glu Asn Lys Ala Cys Leu
130 135 140

Lys Arg Thr Pro Thr Glu Val Trp Arg Asp Ser Arg Asn Pro Tyr Asp
145 150 155 160

Arg Lys Arg Gln Asp Lys Ala Pro Val Gly Leu Lys Asn Val Gly Asn
165 170 175

Thr Cys Trp Phe Ser Ala Val Ile Gln Ser Leu Phe Asn Leu Leu Glu
180 185 190

Phe Arg Arg Leu Val Leu Asn Tyr Lys Pro Pro Ser Asn Ala Gln Asp
195 200 205

Leu Pro Arg Asn Gln Lys Glu His Arg Asn Leu Pro Phe Met Arg Glu
210 215 220

Leu Arg Tyr Leu Phe Ala Leu Leu Val Gly Thr Lys Arg Lys Tyr Val
225 230 235 240

Asp Pro Ser Arg Ala Val Glu Ile Leu Lys Asp Ala Phe Lys Ser Asn
245 250 255

Asp Ser Gln Gln Gln Asp Val Ser Glu Phe Thr His Lys Leu Leu Asp
260 265 270

Trp Leu Glu Asp Ala Phe Gln Met Lys Ala Glu Glu Glu Thr Asp Glu
275 280 285

Glu Lys Pro Lys Asn Pro Met Val Glu Leu Phe Tyr Gly Arg Phe Leu
290 295 300

Ala Val Gly Val Leu Glu Gly Lys Lys Phe Glu Asn Thr Glu Met Phe
305 310 315 320

Gly Gln Tyr Pro Leu Gln Val Asn Gly Phe Lys Asp Leu His Glu Cys

325 330 335

Leu Glu Ala Ala Met Ile Glu Gly Glu Ile Glu Ser Leu His Ser Glu
340 345 350

Asn Ser Gly Lys Ser Gly Gln Glu His Trp Phe Thr Glu Leu Pro Pro
355 360 365

Val Leu Thr Phe Glu Leu Ser Arg Phe Glu Phe Asn Gln Ala Leu Gly
370 375 380

Arg Pro Glu Lys Ile His Asn Lys Leu Glu Phe Pro Gln Val Leu Tyr
385 390 395 400

Leu Asp Arg Tyr Met His Arg Asn Arg Glu Ile Thr Arg Ile Lys Arg
405 410 415

Glu Glu Ile Lys Arg Leu Lys Asp Tyr Leu Thr Val Leu Gln Gln Arg
420 425 430

Leu Glu Arg Tyr Leu Ser Tyr Gly Ser Gly Pro Lys Arg Phe Pro Leu
435 440 445

Val Asp Val Leu Gln Tyr Ala Leu Glu Phe Ala Ser Ser Lys Pro Val
450 455 460

Cys Thr Ser Pro Val Asp Asp Ile Asp Ala Ser Ser Pro Pro Ser Gly
465 470 475 480

Ser Ile Pro Ser Gln Thr Leu Pro Ser Thr Thr Glu Gln Gln Gly Ala
485 490 495

Leu Ser Ser Glu Leu Pro Ser Thr Ser Pro Ser Ser Val Ala Ala Ile
500 505 510

Ser Ser Arg Ser Val Ile His Lys Pro Phe Thr Gln Ser Arg Ile Pro
515 520 525

Pro Asp Leu Pro Met His Pro Ala Pro Arg His Ile Thr Glu Glu Lys
530 535 540

Leu Ser Val Leu Glu Ser Cys Leu His Arg Trp Arg Thr Glu Ile Glu
545 550 555 560

Asn Asp Thr Arg Asp Leu Gln Glu Ser Ile Ser Arg Ile His Arg Thr
565 570 575

Ile Glu Leu Met Tyr Ser Asp Lys Ser Met Ile Gln Val Pro Tyr Arg

```
            580                 585                 590

Leu His Ala Val Leu Val His Glu Gly Gln Ala Asn Ala Gly His Tyr
        595                 600                 605

Trp Ala Tyr Ile Phe Asp His Arg Glu Ser Arg Trp Met Lys Tyr Asn
    610                 615                 620

Asp Ile Ala Val Thr Lys Ser Ser Trp Glu Glu Leu Val Arg Asp Ser
625                 630                 635                 640

Phe Gly Gly Tyr Arg Asn Ala Ser Ala Tyr Cys Leu Met Tyr Ile Asn
                645                 650                 655

Asp Lys Ala Gln Phe Leu Ile Gln Glu Glu Phe Asn Lys Glu Thr Gly
            660                 665                 670

Gln Pro Leu Val Gly Ile Glu Thr Leu Pro Pro Asp Leu Arg Asp Phe
        675                 680                 685

Val Glu Glu Asp Asn Gln Arg Phe Glu Lys Glu Leu Glu Glu Trp Asp
    690                 695                 700

Ala Gln Leu Ala Gln Lys Ala Leu Gln Glu Lys Leu Leu Ala Ser Gln
705                 710                 715                 720

Lys Leu Arg Glu Ser Glu Thr Ser Val Thr Thr Ala Gln Ala Ala Gly
                725                 730                 735

Asp Pro Glu Tyr Leu Glu Gln Pro Ser Arg Ser Asp Phe Ser Lys His
            740                 745                 750

Leu Lys Glu Glu Thr Ile Gln Ile Ile Thr Lys Ala Ser His Glu His
        755                 760                 765

Glu Asp Lys Ser Pro Glu Thr Val Leu Gln Ser Ile Met Met Thr Pro
    770                 775                 780

Asn Met Gln Gly Ile Ile Met Ala Ile Gly Lys Ser Arg Ser Val Tyr
785                 790                 795                 800

Asp Arg Cys Gly Pro Glu Ala Gly Phe Phe Lys Ala Ile Lys Leu Glu
                805                 810                 815

Tyr Ala Arg Leu Val Lys Leu Ala Gln Glu Asp Thr Pro Pro Glu Thr
            820                 825                 830

Asp Tyr Arg Leu His His Val Val Val Tyr Phe Ile Gln Asn Gln Ala
```

```
        835                 840                 845

Pro Lys Lys Ile Ile Glu Lys Thr Leu Leu Glu Gln Phe Gly Asp Arg
    850                 855                 860

Asn Leu Ser Phe Asp Glu Arg Cys His Asn Ile Met Lys Val Ala Gln
865                 870                 875                 880

Ala Lys Leu Glu Met Ile Lys Pro Glu Glu Val Asn Leu Glu Glu Tyr
                885                 890                 895

Glu Glu Trp His Gln Asp Tyr Arg Lys Phe Arg Glu Thr Thr Met Tyr
            900                 905                 910

Leu Ile Ile Gly Leu Glu Asn Phe Gln Arg Glu Ser Tyr Ile Asp Ser
        915                 920                 925

Leu Leu Phe Leu Ile Cys Ala Tyr Gln Asn Asn Lys Glu Leu Leu Ser
    930                 935                 940

Lys Gly Leu Tyr Arg Gly His Asp Glu Glu Leu Ile Ser His Tyr Arg
945                 950                 955                 960

Arg Glu Cys Leu Leu Lys Leu Asn Glu Gln Ala Ala Glu Leu Phe Glu
                965                 970                 975

Ser Gly Glu Asp Arg Glu Val Asn Asn Gly Leu Ile Ile Met Asn Glu
            980                 985                 990

Phe Ile Val Pro Phe Leu Pro Leu Leu Leu Val Asp Glu Met Glu Glu
        995                1000                1005

Lys Asp Ile Leu Ala Val Glu Asp Met Arg Asn Arg Trp Cys Ser Tyr
   1010                1015                1020

Leu Gly Gln Glu Met Glu Pro His Leu Gln Glu Lys Leu Thr Asp Phe
1025                1030                1035                1040

Leu Pro Lys Leu Leu Asp Cys Ser Met Glu Ile Lys Ser Phe His Glu
                1045                1050                1055

Pro Pro Lys Leu Pro Ser Tyr Ser Thr His Glu Leu Cys Glu Arg Phe
            1060                1065                1070

Ala Arg Ile Met Leu Ser Leu Ser Arg Thr Pro Ala Asp Gly Arg
        1075                1080                1085
```

<210> 6
<211> 3803
<212> DNA
<213> Homo sapiens

<400> 6
acagtcggcg tttcgccgcc tgcccgcggt gcccgcgcac gccggccgcc atcgccttcg 60
cgcctggctg gcgggggcgc tgtcctccca ggccgtccgc gccgctccct ggagctcggc 120
ggagcgcggc agccagggcc ggcggaggcg cgaggagccg ggcgccaccg ccgccgccgc 180
cgccgccgcc gcgggggcca tgaccgtgga gcagaacgtg ctgcagcaga gcgcggcgca 240
gaagcaccag cagacgtttt tgaatcaact gagagaaatt acggggatta atgacaccca 300
gatactacag caagccttga aggatagtaa tggaaacttg gaattagcag tggctttcct 360
tactgcgaag aatgctaaga cccctcagca ggaggagaca acttactacc aaacagcact 420
tcctggcaat gatagataca tcagtgtggg aagccaagca gatacaaatg tgattgatct 480
cactggagat gataaagatg atcttcagag agcaattgcc ttgagtttgg ccgaatcaaa 540
cagggcattc agggagactg gaataactga tgaggaacaa gccattagca gagttcttga 600
agccagcata gcagagaata aagcatgttt gaagaggaca cctacagaag tttggaggga 660
ttctcgaaac ccttatgata gaaaaagaca ggacaaagct cccgttgggc taaagaatgt 720
tggcaatact tgttggttta gtgctgttat tcagtcatta tttaatcttt tggaatttag 780
aagattagtt ctgaattaca agcctccatc aaatgctcaa gatttacccc gaaaccaaaa 840
ggaacatcgg aatttgcctt ttatgcgtga gctgaggtat ctatttgcac ttcttgttgg 900
taccaaaagg aagtatgttg atccatcaag agcagttgaa attcttaagg atgctttcaa 960
atcaaatgac tcacagcagc aagatgtgag tgagtttaca cacaaattat tagattggtt 1020
agaagatgcc ttccaaatga aagctgaaga ggagacggat gaagagaagc caaagaaccc 1080
catggtagag ttgttctatg gcagattcct ggctgtggga gtacttgaag gtaaaaaatt 1140
tgaaaacact gaaatgtttg gtcagtaccc acttcaggtc aatgggttca aagatctgca 1200
tgagtgccta gaagctgcaa tgattgaagg agaaattgag tctttacatt cagagaattc 1260
aggaaaatca ggccaagagc attggtttac tgaattacca cctgtgttaa catttgaatt 1320
gtcaagattt gaatttaatc aggcattggg aagaccagaa aaaattcaca acaaattaga 1380
atttccccaa gttttatatt tggacagata catgcacaga aacagagaaa taacaagaat 1440
taagagggaa gagatcaaga gactgaaaga ttacctcacg gtattacaac aaaggctaga 1500
aagatattta agctatggtt ccggtcccaa acgattcccc ttggtagatg ttcttcagta 1560
tgcattggaa tttgcctcaa gtaaacctgt ttgcacttct cctgttgacg atattgacgc 1620
tagttcccca cctagtggtt ccataccatc acagacatta ccaagcacaa cagaacaaca 1680
gggagcccta tcttcagaac tgccaagcac atcaccttca tcagttgctg ccatttcatc 1740
gagatcagta atacacaaac catttactca gtcccggata cctccagatt tgcccatgca 1800
tccggcacca aggcacataa cggaggaaga actttctgtg ctggaaagtt gtttacatcg 1860
ctggaggaca gaaatagaaa atgacaccag agatttgcag gaaagcatat ccagaatcca 1920
tcgaacaatt gaattaatgt actctgacaa atctatgata caagttcctt atcgattaca 1980
tgccgtttta gttcacgaag gccaagctaa tgctgggcac tactgggcat atatttttga 2040
tcatcgtgaa agcagatgga tgaagtacaa tgatattgct gtgacaaaat catcatggga 2100
agagctagtg agggactctt ttggtggtta tagaaatgcc agtgcatact gtttaatgta 2160
cataaatgat aaggcacagt tcctaataca agaggagttt aataaagaaa ctgggcagcc 2220
ccttgttggt atagaaacat taccaccgga tttgagagat tttgttgagg aagacaacca 2280
acgatttgaa aaagaactag aagaatggga tgcacaactt gcccagaaag ctttgcagga 2340
aaagctttta gcgtctcaga aattgagaga gtcagagact tctgtgacaa cagcacaagc 2400
agcaggagac ccagaatatc tagagcagcc atcaagaagt gatttctcaa agcacttgaa 2460

agaagaaact attcaaataa ttaccaaggc atcacatgag catgaagata aaagtcctga 2520
aacagttttg cagtcggcaa ttaagttgga atatgcaagg ttggttaagt tggcccaaga 2580
agacacccca ccagaaaccg attatcgttt acatcatgta gtggtctact ttatccagaa 2640
ccaggcacca aagaaaatta ttgagaaaac attactagaa caatttggag atagaaattt 2700
gagttttgat gaaaggtgtc acaacataat gaaagttgct caagccaaac tggaaatgat 2760
aaaacctgaa gaagtaaact tggaggaata tgaggagtgg catcaggatt ataggaaatt 2820
cagggaaaca actatgtatc tcataattgg gctagaaaat tttcaaagag aaagttatat 2880
agattccttg ctgttcctca tctgtgctta tcagaataac aaagaactct tgtctaaagg 2940
cttatacaga ggacatgatg aagaattgat atcacattat agaagagaat gtttgctaaa 3000
attaaatgag caagccgcag aactcttcga atctggagag gatcgagaag taaacaatgg 3060
tttgattatc atgaatgagt ttattgtccc atttttgcca ttattactgg tggatgaaat 3120
ggaagaaaag gatatactag ctgtagaaga tatgagaaat cgatggtgtt cctaccttgg 3180
tcaagaaatg gaaccacacc tccaagaaaa gctgacagat tttttgccaa aactgcttga 3240
ttgttctatg gagattaaaa gtttccatga gccaccgaag ttaccttcat attccacgca 3300
tgaactctgt gagcgatttg cccgaatcat gttgtccctc agtcgaactc ctgctgatgg 3360
aagataaact gcacactttc cctgaacaca ctgtataaac tctttttagt tcttaaccct 3420
tgccttcctg tcacagggtt tgcttgttgc tgctatagtt tttaactttt ttttatttta 3480
ataactgcaa aagacaaaat gactatacag actttagtca gactgcagac aataaagctg 3540
aaaatcgcat ggcgctcaga cattttaacc ggaactgatg tataatcaca aatctaattg 3600
attttattat ggcaaaacta tgcttttgcc accttcctgt tgcagtatta ctttgctttt 3660
atcttttctt tctcaacagc tttccattca gtctggatcc ttccatgact acagccattt 3720
aagtgttcag cactgtgtac gatacataat atttggtagc ttgtaaatga aataaagaat 3780
aaagttttat ttatggctac cta 3803

<210> 7
<211> 3169
<212> DNA
<213> Homo sapiens

<400> 7
catgaccgtg gagcagaacg tgctgcagca gagcgcggcg cagaagcacc agcagacgtt 60
tttgaatcaa ctgagagaaa ttacggggat taatgacacc cagatactac agcaagcctt 120
gaaggatagt aatggaaact tggaattagc agtggctttc cttactgcga agaatgctaa 180
gacccctcag caggaggaga caacttacta ccaaacagca cttcctggca atgatagata 240
catcagtgtg ggaagccaag cagatacaaa tgtgattgat ctcactggag atgataaaga 300
tgatcttcag agagcaattg ccttgagttt ggccgaatca aacagggcat tcagggagac 360
tggaataact gatgaggaac aagccattag cagagttctt gaagccagca tagcagagaa 420
taaagcatgt ttgaagagga cacctacaga agtttggagg gattctcgaa acccttatga 480
tagaaaaaga caggacaaag ctcccgttgg gctaaagaat gttggcaata cttgttggtt 540
tagtgctgtt attcagtcat tatttaatct tttggaattt agaagattag ttctgaatta 600
caagcctcca tcaaatgctc aagatttacc ccgaaaccaa aaggaacatc ggaatttgcc 660
ttttatgcgt gagctgaggt atctatttgc acttcttgtt ggtaccaaaa ggaagtatgt 720
tgatccatca agagcagttg aaattcttaa ggatgctttc aaatcaaatg actcacagca 780
gcaagatgtg agtgagttta cacacaaatt attagattgg ttagaagatg ccttccaaat 840
gaaagctgaa gaggagacgg atgaagagaa gccaaagaac cccatggtag agttgttcta 900
tggcagattc ctggctgtgg gagtacttga aggtaaaaaa tttgaaaaca ctgaaatgtt 960
tggtcagtac ccacttcagg tcaatgggtt caaagatctg catgagtgcc tagaagctgc 1020

aatgattgaa ggagaaattg agtctttaca ttcagagaat tcaggaaaat caggccaaga 1080
gcattggttt actgaattac cacctgtgtt aacatttgaa ttgtcaagat ttgaatttaa 1140
tcaggcattg ggaagaccag aaaaaattca caacaaatta gaatttcccc aagttttata 1200
tttggacaga tacatgcaca gaaacagaga aataacaaga attaagaggg aagagatcaa 1260
gagactgaaa gattacctca cggtattaca acaaaggcta gaaagatatt taagctatgg 1320
ttccggtccc aaacgattcc ccttggtaga tgttcttcag tatgcattgg aatttgcctc 1380
aagtaaacct gtttgcactt ctcctgttga cgatattgac gctagttccc cacctagtgg 1440
ttccatacca tcacagacat taccaagcac aacagaacaa cagggagccc tatcttcaga 1500
actgccaagc acatcacctt catcagttgc tgccatttca tcgagatcag taatacacaa 1560
accatttact cagtcccgga tacctccaga tttgcccatg catccggcac caaggcacat 1620
aacggaggaa gaactttctg tgctggaaag ttgtttacat cgctggagga cagaaataga 1680
aaatgacacc agagatttgc aggaaagcat atccagaatc catcgaacaa ttgaattaat 1740
gtactctgac aaatctatga tacaagttcc ttatcgatta catgccgttt tagttcacga 1800
aggccaagct aatgctgggc actactgggc atatattttt gatcatcgtg aaagcagatg 1860
gatgaagtac aatgatattg ctgtgacaaa atcatcatgg gaagagctag tgagggactc 1920
ttttggtggt tatagaaatg ccagtgcata ctgtttaatg tacataaatg ataaggcaca 1980
gttcctaata caagaggagt ttaataaaga aactgggcag ccccttgttg gtatagaaac 2040
attaccaccg gatttgagag attttgttga ggaagacaac caacgatttg aaaaagaact 2100
agaagaatgg gatgcacaac ttgcccagaa agctttgcag gaaaagcttt tagcgtctca 2160
gaaattgaga gagtcagaga cttctgtgac aacagcacaa gcagcaggag acccagaata 2220
tctagagcag ccatcaagaa gtgatttctc aaagcacttg aaagaagaaa ctattcaaat 2280
aattaccaag gcatcacatg agcatgaaga taaaagtcct gaaacagttt tgcagtcggc 2340
aattaagttg gaatatgcaa ggttggttaa gttggcccaa gaagacaccc caccagaaac 2400
cgattatcgt ttacatcatg tagtggtcta ctttatccag aaccaggcac caaagaaaat 2460
tattgagaaa acattactag aacaatttgg agatagaaat ttgagttttg atgaaaggtg 2520
tcacaacata atgaaagttg ctcaagccaa actggaaatg ataaaacctg aagaagtaaa 2580
cttggaggaa tatgaggagt ggcatcagga ttataggaaa ttcagggaaa caactatgta 2640
tctcataatt gggctagaaa attttcaaag agaaagttat atagattcct tgctgttcct 2700
catctgtgct tatcagaata acaaagaact cttgtctaaa ggcttataca gaggacatga 2760
tgaagaattg atatcacatt atagaagaga atgtttgcta aaattaaatg agcaagccgc 2820
agaactcttc gaatctggag aggatcgaga agtaaacaat ggtttgatta tcatgaatga 2880
gtttattgtc ccatttttgc cattattact ggtggatgaa atggaagaaa aggatatact 2940
agctgtagaa gatatgagaa atcgatggtg ttcctacctt ggtcaagaaa tggaaccaca 3000
cctccaagaa aagctgacag atttttgcc aaaactgctt gattgttcta tggagattaa 3060
aagtttccat gagccaccga agttaccttc atattccacg catgaactct gtgagcgatt 3120
tgcccgaatc atgttgtccc tcagtcgaac tcctgctgat ggaagataa 3169